



**Análise de clusters espaço-temporais
de anomalias congénitas em Portugal**

Afonso Manuel Lohmann da Silva Rodrigues

Mestrado em Bioestatística

Relatório de Estágio orientado por:
Professora Doutora Marília Antunes
Mestre Paula Braz

Resumo

Este trabalho foi realizado no âmbito da Unidade Curricular de Dissertação/Estágio/Projeto, do mestrado em Bioestatística da Faculdade de Ciências da Universidade de Lisboa, e teve como principal objetivo a criação ou adaptação de uma ferramenta computacional que permitisse ao Registo Nacional de Anomalias Congénitas (RENAC), do Instituto Nacional de Saúde Doutor Ricardo Jorge, realizar a análise de *clusters* espaço-temporais de anomalias congénitas (AC).

Os resultados da análise de *clusters* neste contexto podem ajudar a detetar locais e intervalos temporais em que se verifica um aumento do número de casos de AC. Com essa informação, torna-se possível atuar na prevenção junto das populações ou iniciarem-se estudos de relação entre AC específicas e fatores teratogénicos, i.e., fatores passíveis de provocar AC, ao perturbarem o desenvolvimento embrionário ou fetal.

A ferramenta criada tem como principal propósito a facilitação da utilização do programa de livre acesso SaTScanTM, que efetua análises espaço-temporais de diversos eventos com base em modelos estatísticos. A metodologia utilizada, denominada estatística de varrimento de janela variável, é análoga à utilizada pela rede europeia de registos, o *European Surveillance of Congenital Anomalies* (EUROCAT), com a principal diferença de neste último serem pesquisados apenas *clusters* temporais.

No código da ferramenta estão contidos os procedimentos que permitem a adaptação automática da base de dados do RENAC, de dados fornecidos pelo Instituto Nacional de Estatística referentes aos nascimentos e das coordenadas geográficas dos concelhos de Portugal ao programa, e a possibilidade de definir os parâmetros das análises a realizar sem ser necessário o conhecimento pormenorizado do funcionamento do SaTScanTM. A fim de testar o funcionamento desta ferramenta foram realizadas diversas análises de *clusters* espaço-temporais de AC em Portugal cujos resultados serão apresentados e discutidos.

Esta ferramenta permite a realização de um número elevado de análises num breve intervalo de tempo e aparenta ter utilidade na pesquisa de *clusters* espaço-temporais de AC em Portugal, ainda que os resultados que dela advêm devam ser cuidadosamente interpretados e revistos considerando as características específicas dos dados do RENAC.

Palavras-chave: anomalias congénitas, RENAC, teratogénico, *clusters*, SaTScanTM

Abstract

This work was carried out in the context of the Faculdade de Ciências da Universidade de Lisboa's Master's in Biostatistics' Curricular Unit of Dissertation/Internship/Project. It's main goal was to create or adapt a computational tool that would allow the National Registry of Congenital Anomalies, Registo Nacional de Anomalias Congénitas (RENAC), of the National Health Institute, Instituto Nacional de Saúde Doutor Ricardo Jorge, to execute the analysis of spatio-temporal clusters of congenital anomalies (CA).

In these circumstances, the results of clusters analysis can help to detect areas and time periods in which it can be verified an increase in the number of cases of CA. With this information, it is possible to act on the prevention on the populations or begin connexion studies between specific CA and teratogenic factors, i.e., factors which can cause CA by disturbing the embryonic or fetal development.

The created tool has the chief purpose of facilitating the use of the freeware SaTScanTM, which runs spatio-temporal analysis of various events based on statistical models. The used methodology, referred to as the variable window scan statistic, is similar to the one used by the european registry network, the European Surveillance of Congenital Anomalies (EUROCAT), with the main difference that, in the latter, the cluster analysis is solely temporal.

In the tool's script are the procedures that allow the automatic filtering of the RENAC's database, the birth data from the National Statistics Institute and the geographic coordinates of Portugal's municipalities to fit the SaTScanTM's standards, as well as the possibility of defining the parameters to be used in the analysis without having in depth knowledge of the program. In order to test the tool's behaviour, a series of spatio-temporal cluster analysis of CA cases were run, whose results will be presented and reviewed.

This tool allows for a multitude of analysis to be performed in a short period of time and appears to be useful in the research of spatio-temporal clusters of CA in Portugal, although the results should be carefully interpreted and reviewed, considering the specific characteristics of RENAC's data.

Keywords: congenital anomalies, RENAC, teratogenic, clusters, SaTScanTM

Índice

	Pág.
Resumo	i
Abstract	ii
Lista de quadros	v
Lista de figuras	vii
Lista de abreviaturas e siglas	viii
1 Enquadramento teórico e objetivos	1
1.1 Anomalias congénitas	1
1.2 O RENAC	2
1.3 O EUROCAT	3
1.4 <i>Clusters</i> e a sua importância	4
1.5 A metodologia de identificação de <i>clusters</i> temporais do EUROCAT	6
1.6 A estatística de varrimento espacial e espaço-temporal	11
1.7 Objetivos	14
1.7.1 Objetivo principal	14
1.7.2 Objetivos secundários	14
2 Materiais e métodos	15
2.1 Dados utilizados	15
2.1.1 Dados das anomalias congénitas	15
2.1.2 Dados populacionais	16
2.1.3 Coordenadas dos concelhos de Portugal	17
2.2 Metodologia	19
2.3 Funcionamento da ferramenta	27
2.3.1 A inserção dos dados	27
2.3.2 Procedimentos incluídos na ferramenta	28

2.3.3	Interpretação do <i>output</i>	29
3	Resultados e conclusões	33
3.1	Resultados de <i>clusters</i> espácio-temporais de AC para o ano de 2013	34
3.2	Resultados de <i>clusters</i> espácio-temporais de AC do aparelho circulatório (Q20-Q28)	42
3.3	Conclusões	56
	Referências bibliográficas	57
	Anexos	61
Anexo 1:	Instrumento de notificação do RENAC	61
Anexo 2:	Variáveis presentes na base de dados do RENAC	64
Anexo 3:	<i>Script</i> integral da ferramenta	71
Anexo 4:	Instruções de utilização da ferramenta de análise de <i>clusters</i> espácio-temporais .	79

Lista de quadros

2.1	Codificação das AC por grandes grupos, de acordo com o capítulo XVII da CID-10, utilizada pelo RENAC entre 2000 e 2013.	16
2.2	Codificação das AC não incluídas no capítulo XVII da CID-10, utilizada pelo RENAC entre 2000 e 2013, e o grupo do mesmo capítulo em que as mesmas foram inseridas.	16
3.1	As AC analisadas e o seu código segundo a CID-10.	33
3.2	Resultados da análise de <i>clusters</i> espácio-temporais de todos os casos de AC reportados ao RENAC ocorridos em 2013 em Portugal continental.	35
3.3	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de AC do SNC (Q00-Q07) reportados ao RENAC ocorridos em 2013 em Portugal continental.	36
3.4	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de defeitos do tubo neural (Q00,Q01,Q05) reportados ao RENAC ocorridos em 2013 em Portugal continental.	36
3.5	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de microcefalia (Q02) reportados ao RENAC ocorridos em 2013 em Portugal continental.	37
3.6	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de espinha bífida (Q05) reportados ao RENAC ocorridos em 2013 em Portugal continental.	37
3.7	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de AC do olho, ouvido, face e pescoço (Q10-Q18) reportados ao RENAC ocorridos em 2013 em Portugal continental.	38
3.8	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de coarctação da aorta (Q25.1) reportados ao RENAC ocorridos em 2013 em Portugal continental.	38
3.9	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de AC do aparelho respiratório (Q30-Q34) reportados ao RENAC ocorridos em 2013 em Portugal continental.	39
3.10	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de AC do sistema digestivo não incluídas nos restantes grandes grupos (Q38-Q45) reportados ao RENAC ocorridos em 2013 em Portugal continental.	39
3.11	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013 em Portugal continental.	40
3.12	Resultados da análise de <i>clusters</i> espácio-temporais dos casos de AC do sistema músculo-esquelético (Q65-Q79) reportados ao RENAC ocorridos em 2013 em Portugal continental.	41

3.13	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de sindactilia (Q70) reportados ao RENAC ocorridos em 2013 em Portugal continental.	41
3.14	Resultados da análise de <i>clusters</i> espaço-temporais dos casos com outras AC não incluídas em qualquer outro grande grupo (Q80-Q89) reportados ao RENAC ocorridos em 2013 em Portugal continental.	42
3.15	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2000 em Portugal continental.	43
3.16	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2001 em Portugal continental.	44
3.17	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2002 em Portugal continental.	45
3.18	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2003 em Portugal continental.	46
3.19	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2004 em Portugal continental.	47
3.20	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2005 em Portugal continental.	48
3.21	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2007 em Portugal continental.	49
3.22	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2008 em Portugal continental.	50
3.23	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2009 em Portugal continental.	51
3.24	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2010 em Portugal continental.	52
3.25	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2011 em Portugal continental.	54
3.26	Resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2012 em Portugal continental.	55

Lista de figuras

1.1	Representação gráfica da equação (1.3)	8
2.1	Exemplo do ficheiro com os valores totais de nascimentos para cada concelho de Portugal, por mês, referentes ao ano 2000.	17
2.2	Excerto do ficheiro com os dados geográficos para os municípios de Portugal continental.	18
2.3	Exemplo da matriz de dados dos casos como vista na interface do R	20
2.4	Exemplo da matriz de dados dos controlos como vista na interface do R	21
2.5	Exemplo da matriz de dados com as informações geográficas como vista na interface do R	21
2.6	Colagem da introdução sequencial dos vários parâmetros durante o uso da ferramenta com exemplos de resposta.	28
2.7	Porção inicial do ficheiro com os resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013 em Portugal continental.	32
A.1	Exemplo do ficheiro com os valores totais de nascimentos para cada concelho de Portugal, por mês, referentes ao ano 2000.	81
A.2	Imagem de uma pasta com todos os ficheiros prontos a ser lidos pela ferramenta. . .	82
A.3	(Em cima) imagem da primeira pasta criada, cujo nome indica os anos do período considerado para a análise e que contém a segunda pasta; (em baixo) imagem da segunda pasta criada, no interior da primeira, cujo nome indica as AC escolhidas para a análise.	84
A.4	Porção inicial do ficheiro com os resultados da análise de <i>clusters</i> espaço-temporais dos casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013 em Portugal continental.	87

Lista de abreviaturas e siglas

AC: Anomalia(s) congénita(s)

CDC: *Centers for Disease Control and Prevention*

CID-10: 10^a Revisão da Classificação Internacional de Doenças

CSV: *Comma Separated Values*

DEP: Departamento de Epidemiologia

EDMP: *EUROCAT Data Management Program*

EUROCAT: *European Surveillance of Congenital Anomalies*

GIS: *Geographic Information System* (Sistema de Informação Geográfica)

GLRT: *Generalized Likelihood Ratio Test*

IDH: Índice de Desenvolvimento Humano

IG: Idade Gestacional

IMG: Interrupção Médica da Gravidez

INE: Instituto Nacional de Estatística

INSA: Instituto Nacional de Saúde Doutor Ricardo Jorge

NV: Nado-vivo

OMS: Organização Mundial de Saúde

RENAC: Registo Nacional de Anomalias Congénitas

RIOS: Redes de Informação e Observação em Saúde

SNC: Sistema Nervoso Central

UNDP: *United Nations Development Programme*

WHO: *World Health Organization*

Capítulo 1

Enquadramento teórico e objetivos

1.1 Anomalias congénitas

Anomalias congénitas (AC) definem-se como alterações funcionais ou estruturais que afetam um indivíduo e ocorrem durante a gravidez, podendo ser identificadas durante a gestação, ao nascer ou, por vezes, ao longo da vida (WHO – World Health Organization, 2016). Embora cerca de 50% dos casos de AC tenham etiologia desconhecida, estudos apontam para que condições genéticas, ou exposições teratogénicas ¹, tais como, condições ambientais, comportamentos maternos, entre outros, possam provocar AC através da perturbação do desenvolvimento embrionário (WHO, 2016). Nos países desenvolvidos a prevalência das AC *major*, i.e., defeitos estruturais com relevância médica ou estética, é de aproximadamente 2,5% (Dolk, Loane & Garne, 2010; CDC – Centers for Disease Control and Prevention, 2016). As AC podem ser isoladas – cerca de 70% dos casos – ou múltiplas, sendo que as múltiplas se desenvolvem independentemente, embora a causa tenda a ser comum (CDC, 2016).

De entre as causas de morte durante o período neonatal, que incluem o nascimento prematuro, trauma durante o parto, entre outras, as AC representam também uma proporção importante, não só de mortalidade, mas ainda de morbilidade infantil. O *United Nations Development Programme* (UNDP), a rede global de desenvolvimento das Nações Unidas, utiliza para o cálculo do índice de desenvolvimento humano (IDH) dos países, a esperança de vida à nascença, relacionada de forma clara com a mortalidade infantil, sendo esta última influenciada pela ocorrência de AC (United Nations Development Programme, 2013). As AC tornam-se, por isso, num objeto de estudo importante na área da Saúde Pública: a etiologia, as formas de prevenção, a sua prevalência e distribuição ao longo do espaço e do tempo, formas de deteção mais precisas e precoces, e tratamentos e cuidados possíveis (WHO, 2016; CDC, 2016).

¹Entidades passíveis de perturbar o desenvolvimento embrionário ou fetal e provocar anomalias congénitas (Carvalho et al., 2007a).

Além de serem uma causa relevante de mortalidade infantil, as AC também contribuem para a morbidade e incapacidade dos indivíduos. Como consequência, têm um forte impacto na qualidade de vida do indivíduo, das suas famílias, nos sistemas de saúde, e, por fim, na sociedade. Reveste-se, assim, de toda a importância o conhecimento mais cabal possível da realidade das AC, em Portugal e no mundo. Sendo pertinente analisar a sua distribuição geográfica e temporal, através de ferramentas que permitam detetar eventuais aglomerados de AC. Ao relacionar estes com possíveis exposições teratogénicas, poder-se-á conhecer a sua etiologia e fornecer informações mais precisas aos decisores políticos, auxiliando o desenvolvimento de medidas de prevenção em saúde pública (WHO, 2016; Braz, Machado, Roquette & Dias, 2015).

Enquanto patologia, as AC podem afetar quaisquer sistemas orgânicos do corpo humano, embora com frequências distintas, afetando algumas vezes, mais que um sistema. As anomalias do sistema circulatório e do sistema músculo-esquelético são as que apresentam maior prevalência, tanto em Portugal como na Europa (Braz et al., 2015; Dolk et al., 2015). Algumas das AC têm causas e/ou medidas de prevenção conhecidas. A título ilustrativo, o Síndrome de Down, uma das AC com maior prevalência, provocada pela presença de uma cópia supranumerária, total ou parcial, do cromossoma 21 (Patterson, 2009), tem como fator de risco, entre outros, a idade materna avançada na altura da gravidez; também, os defeitos do tubo neural, incluídas nas AC do sistema nervoso central (SNC), têm como medida de prevenção o aumento do aporte de ácido fólico na mulher (vitamina B9) entre a fase pré-concepcional e o final do primeiro trimestre de gravidez (WHO, 2016; CDC, 2016; Braz et al., 2015).

No nosso país, o estudo dos casos de AC é desenvolvido pelo RENAC, que se apresenta de seguida (Braz et al., 2015).

1.2 O RENAC

O RENAC é um registo nosológico ² de base populacional que recebe notificações, de forma anonimizada, de casos de AC ocorridos em território nacional, sendo um dos instrumentos de monitorização e vigilância em saúde do departamento de epidemiologia (DEP) do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA). São notificados ao RENAC casos com AC diagnosticadas em nados-vivos (NV) detetadas até aos 28 dias de vida (período neonatal), fetos mortos com idade gestacional superior a 22 semanas ou com mais de 500g de peso e fetos submetidos a interrupção médica da gravidez (IMG) de qualquer idade gestacional ou peso. São critérios de exclusão os casos com doenças metabólicas quando não associadas a AC, defeitos resultantes de traumatismos ocorridos durante o parto, e AC *minor* isoladas, i.e., AC sem impacto significativo na saúde neonatal e com consequências estéticas e sociais limitadas (CDC, 2017).

Colaboram com o RENAC, enquanto fontes de informação, maternidades, serviços hospitalares de pediatria e de obstetrícia, serviços de anatomia patológica, serviços de especialidades pediátricas,

²Que lida com o estudo e classificação de doenças (Carvalho et al., 2007b).

e laboratórios de citogenética. O RENAC tem como objetivos: descrever a prevalência de AC e a sua distribuição geográfica face à residência das mães; avaliar a efetividade de medidas de prevenção primária; manter uma base de dados disponível para os serviços participantes, a comunidade científica e o Ministério da Saúde; manter um sistema de vigilância que detete, em tempo útil, variações súbitas que possam evidenciar novas exposições teratogénicas; avaliar o impacto do diagnóstico pré-natal; integrar a rede europeia de registos, *European Surveillance of Congenital Anomalies* (EUROCAT).

A colaboração com o RENAC é voluntária, dependendo por isso da adesão dos centros hospitalares, o que, de acordo com os relatórios publicados, pode conduzir à subnotificação de casos e consequente subestimação das prevalências de AC a nível nacional (Braz et al., 2015; INSA *website*).

O instrumento de notação, ilustrado no Anexo 1, inclui 50 variáveis que recolhem informação sobre o recém-nascido ou o feto, dados da gestação, história pessoal e familiar dos progenitores, exposições teratogénicas da mãe e diagnóstico médico das AC. Esta informação, com dados anonimizados, é enviada para o RENAC, em papel, ou em formato digital, este último através da utilização da plataforma RIOS (Redes de Informação e Observação em Saúde), sediada no INSA, onde fica armazenada a base dados.

1.3 O EUROCAT

O EUROCAT é uma rede de registos europeus de AC, criada em 1979, constituída atualmente por 43 registos, provenientes de 23 países, permitindo a vigilância de mais de 1.7 milhões de nascimentos por ano, i.e., de cerca de 29% dos nascimentos europeus. Estes registos partilham a sua metodologia e critérios de qualidade, mantendo não só um sistema de vigilância europeu de AC, bem como promovendo a sua investigação epidemiológica. Portugal participa no EUROCAT desde 1990, notificando os casos com mães residentes a sul do rio Tejo (Boyd et al., 2011; EUROCAT *website*).

Para além dos objetivos que partilha com o RENAC, o EUROCAT fornece ainda uma rede colaborativa, metodologia e infra-estrutura para a investigação científica. Esta tem como finalidade o estudo das causas, prevenção, tratamento e cuidados a prestar às crianças afetadas por AC. O EUROCAT constitui-se também como um centro de informação para populações, famílias e profissionais de saúde, em assuntos como *clusters* ou aglomerados temporais de AC na Europa, exposições teratogénicas e eventuais fatores de risco, alavancando a criação de registos de AC na Europa, a fim de permitir a recolha de dados standardizados e facilmente comparáveis. Colabora também com a Organização Mundial de Saúde (OMS) na vigilância epidemiológica de AC (Boyd et al., 2011; EUROCAT *website*).

O EUROCAT, constituído após a crise da talidomida, um medicamento utilizado entre aproximadamente 1950 e 1960 para o tratamento de enjoos em grávidas, mais tarde reconhecido como um

potente agente teratogénico (Kim & Scialli, 2011), sempre deu grande relevância ao estudo das exposições fetais durante a gravidez. A metodologia comum aos registos que participam no EUROCAT permite responder ao objetivo de estudar alterações genéticas e as interações entre estas e os fatores ambientais, que podem levar ao desenvolvimento de AC. Também promove a avaliação dos serviços de saúde, em diversas vertentes: na prevenção primária, como é um exemplo a suplementação de ácido fólico para prevenir defeitos do tubo neural; no rastreio pré-natal, com recurso a meios auxiliares de diagnóstico; na reabilitação e no tratamento. Estes dados podem ser usados no sentido de avaliar os cuidados de saúde de cada país, ou região, e dar suporte aos gestores e decisores políticos (Boyd et al., 2011; EUROCAT *website*).

A gravidade da crise da talidomida ilustrou a relevância de detetar excessos da prevalência (*clusters*) de casos de AC específicas a nível nacional e europeu, de modo a intervir o mais precocemente possível para controlar a sua causa. O EUROCAT define um *cluster* como “uma agregação de casos no tempo e/ou no espaço que aparenta ser invulgar”, definição esta adaptada do CDC, e que será a utilizada neste documento de forma permutável com “aglomerado”. Desta forma, o termo é utilizado com um significado estatístico. Neste sentido, “*clustering*” será usado como designação genérica das técnicas estatísticas que permitem a identificação de *clusters*.

Para detetar um número de casos acima do esperado, que pode indiciar a exposição da população a algum fator de risco, o EUROCAT criou uma ferramenta que identifica *clusters* temporais com recurso a estatísticas de varrimento, que realiza testes à mudança da distribuição subjacente aos dados, facilitando a identificação de *clusters* de casos raros (Teljeur, Kelly, Loano, Densem & Dolk, 2015; Dolk et al., 2015).

1.4 *Clusters* e a sua importância

A análise de *clusters* tem vindo a ser utilizada de diversas formas numa multiplicidade de áreas, tais como no *marketing*, para descobrir grupos entre os consumidores e caracterizar esses mesmo grupos; na Biologia, para categorizar genes com funções semelhantes; na Saúde Pública para identificação de *clusters* de casos no espaço e/ou no tempo como parte da monitorização da incidência de doenças.

Motivado pelo aumento no tamanho das bases de dados, o *clustering* é uma área com um desenvolvimento acentuado, contando com contribuições de variadas áreas de estudo, como a estatística, o *marketing* e o *data mining*; neste último, o processo de *clustering* consiste em agrupar um conjunto de objetos em classes, ou *clusters*, de tal modo que aqueles que se encontram no interior de um *cluster* sejam semelhantes entre si e, simultaneamente, diferentes dos que se encontram fora da mesma. Enquanto ramo da estatística, a análise de *clusters* foca-se principalmente na análise baseada na distância (Han & Kamber, 2001).

Existe atualmente um elevado número de algoritmos de *clustering*, sendo a sua classificação difícil e, muitas vezes, um método pode ser incluído em mais que uma categoria. Não obstante, alguns autores

consideram as seguintes categorias (Han & Kamber, 2001), que serão apresentadas de seguida: métodos de partição, métodos hierárquicos, métodos baseados na densidade, métodos baseados em modelos e métodos baseados numa grelha (*grid-based methods*) (Han & Kamber, 2001).

Algoritmos de partição

Os algoritmos de partição (*partitioning*, no original) constroem uma segmentação da base de dados D com n objetos num conjunto de k *clusters*. Tipicamente, começa-se com uma partição inicial de D , seguindo-se uma estratégia de controlo iterativo para otimizar uma função objetivo. Cada *cluster* é representado pelo respetivo centro de gravidade (centroide) ou por um objeto próximo desse mesmo centro. Como consequência, estes algoritmos utilizam um processo com dois passos: determinar k representantes que minimizem a função objetivo e associar cada objeto ao *cluster* que tenha o seu representante mais próximo do mesmo. O segundo passo implica que o resultado da partição pode ser representado por um diagrama de Voronoi (uma partição de um plano em células), cada objeto pertencendo a uma célula de Voronoi. Por conseguinte, os *clusters* encontrados são convexos, o que pode ser considerado restritivo (Ester, Kriegel, Sander & Xu, 1996; Han & Kamber, 2001).

Algoritmos hierárquicos

Os algoritmos hierárquicos têm na sua base uma decomposição sucessiva da base de dados D . Esta decomposição é representada por um dendrograma, um diagrama em árvore, que, de forma iterativa, divide D em subconjuntos cada vez menores até cada subconjunto ser constituído por apenas um objeto. Assim, cada nó da árvore representa um *cluster* de D . O dendrograma pode ser criado de duas formas: a partir das folhas até à raiz, ou de forma inversa, da raiz até às folhas. Temos, como consequência, a abordagem aglomerativa e divisiva, respetivamente. Estes algoritmos requerem uma condição de paragem para indicação de quando o processo de aglomeração ou divisão deve terminar. A título de exemplo, uma condição de paragem comum na abordagem aglomerativa é a distância crítica, D_{min} , entre todos os *clusters*. Daqui advém o problema de definir D_{min} tal que seja um valor pequeno o suficiente para separar todos os *clusters* existentes, mas, em simultâneo, grande o bastante para não permitir a divisão dos mesmos (Ester et al., 1996).

Algoritmos baseados na densidade

Os algoritmos de *clustering* baseados na densidade surgiram devido às restrições dos dois métodos acima expostos. A ideia por detrás destes algoritmos é que os *clusters* continuem a aumentar de tamanho enquanto a densidade, i.e., o número de objetos no interior da sua vizinhança (*neighborhood*), exceder um dado limiar. Ou seja, para cada objeto no interior de um determinado *cluster*, a vizinhança, delimitada por um raio fixo, tem que conter pelo menos um número mínimo de objetos. Este método permite então filtrar o ruído, pontos que não pertencem a qualquer *cluster* e potenciais

outliers, bem como encontrar *clusters* com uma forma arbitrária (Han & Kamber, 2001).

Algoritmos baseados em modelos

Um algoritmo de *clustering* baseado em modelos pode encontrar *clusters* ao construir uma função densidade de probabilidade que reflita a distribuição espacial dos dados, e pode ainda permitir a determinação automática do número de *clusters* com base na estatística, tendo os *outliers* e/ou ruído em consideração (Han & Kamber, 2001).

Algoritmos baseados em grelha

Por último, os métodos baseados em grelha assentam na construção de uma grelha multidimensional cobrindo o espaço onde se encontram os objetos, induzindo a criação de um número de células que dependerá da granularidade desejada. Todos os objetos pertencentes a uma mesma célula são representados por um único ponto, com características definidas à custa de estatísticas dos referidos objetos (por exemplo, o centroide). A redução da dimensionalidade daí decorrente depende da quantidade de objetos na base de dados original e de quão fina é a grelha. Desta forma, todas as operações de *clustering* são realizadas nessa estrutura em grelha, sobre os representantes das células, o que diminui o tempo de processamento de forma inversamente proporcional ao quão fina é a grelha, tornando-se este o fator mais importante no tempo de processamento, e não o número de objetos na base de dados (Han & Kamber, 2001).

1.5 A metodologia de identificação de *clusters* temporais do EUROCAT

O EUROCAT realiza uma monitorização estatística central que deteta *clusters* temporais dentro de cada área de registo (um país ou uma região). O *software* EUROCAT *Data Management Program* (EDMP) permite ainda a cada área de registo detetar *clusters* temporais localmente dentro de subáreas da mesma. A deteção de *clusters* é realizada com base no teste descrito por Naus e Nagarwalla que é exposto de seguida (EUROCAT Statistical Monitoring Protocol, 2009).

Estatística de Varrimento

Sejam X_1, \dots, X_N variáveis aleatórias (v.a.) independentes e identicamente distribuídas (i.i.d.) que representam os tempos de ocorrência de N eventos num intervalo $]0, T]$. Testar-se-á a hipótese nula que afirma que os eventos são distribuídos de forma uniforme contra a hipótese alternativa que considera que os eventos formam um aglomerado nalgum subintervalo de $]0, T]$. Sem perda de

generalidade, faça-se $T = 1$ (Nagarawalla, 1996; Naus, 1965).

Admitindo X_i , $i = 1, \dots, N$, v.a. i.i.d. uniformes $(0,1)$, Naus (1965) descreveu a probabilidade de existir um subintervalo de $]0, 1]$ de comprimento fixo $d < 1$ que contém n ou mais dos N eventos. É aqui que surge a estatística de varrimento simples, ou de janela fixa. A estatística de varrimento simples, N_d é então o número máximo de eventos, ou pontos, observados num intervalo de duração fixa d , tomando este intervalo o nome de janela de varrimento, que se move ao longo do intervalo $]0, 1]$ de forma contínua (Nagarawalla, 1996; Naus, 1965).

Nas suas aplicações epidemiológicas surge um problema com o uso desta estatística. Ao fixar o valor de d é necessário determinar o tamanho potencial dos *clusters* antes de analisar os dados a fim de evitar enviesar a análise e manter a significância do teste. Se um *cluster* real abarcar um intervalo de duração muito inferior ou muito superior, a potência do teste poderá ser afetada. Não obstante, o teste baseado na estatística de varrimento simples é o *Generalized Likelihood Ratio Test* (GLRT), para testar a hipótese nula que afirma que a distribuição dos eventos ao longo do intervalo $]0, 1]$ é uniforme contra uma hipótese alternativa que sugere alguma forma de aglomeração dos eventos. O tipo de aglomeração especificado nesta classe de alternativas compreende a existência de um subintervalo de $]0, 1]$ de amplitude (duração) fixa d , tal que, ao longo deste subintervalo, o risco de ser um caso, ou evento, é aumentado por influência de um fator constante desconhecido, relativamente ao risco fora do mesmo. Posteriormente será introduzida a estatística de varrimento com janela variável (Nagarawalla, 1996; Naus, 1965).

Seja (X_1, \dots, X_N) uma amostra aleatória de N pontos com a função densidade de probabilidade

$$f(x) = \begin{cases} a, & \text{se } b \leq x \leq b + d \\ \frac{1-ad}{1-d}, & \text{se } 0 \leq x < b \text{ ou } b + d < x \leq 1 \end{cases}, \quad (1.1)$$

sendo as hipóteses em teste

$$H_0 : a = 1 \text{ vs } H_1 : 1 < a \leq 1/d \quad . \quad (1.2)$$

A hipótese nula afirma que a distribuição do número de eventos é uniforme ao longo do intervalo $]0, 1]$, dentro e fora da janela de varrimento. Por sua vez, a hipótese alternativa indica que no interior da janela de varrimento o número de eventos segue uma distribuição uniforme de valor mais elevado que no seu exterior, garantindo, no entanto, que o seu integral no intervalo $]0, 1]$ possui um valor unitário.

Para ilustrar a equação (1.1), definam-se os parâmetros $a = 1,6$, $b = 0,4$ e $d = 0,4$ e tem-se a função densidade de probabilidade

$$f(x) = \begin{cases} 1,6 & , \quad \text{se } 0,4 \leq x \leq 0,8 \\ 0,6 & , \quad \text{se } 0 \leq x < 0,4 \text{ ou } 0,8 < x \leq 1 \end{cases} \quad (1.3)$$

representada graficamente na figura seguinte.

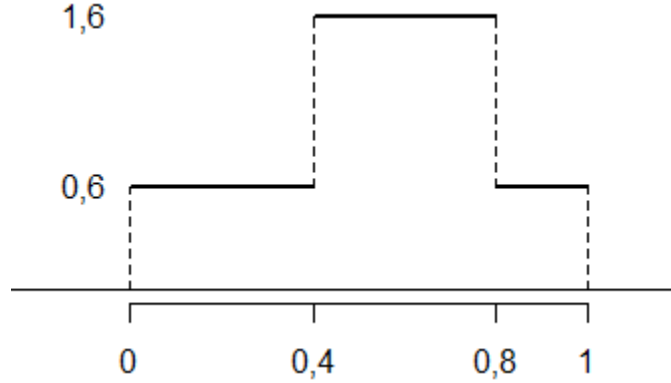


Figura 1.1: Representação gráfica da equação (1.3)

O teste a realizar, como indicado supra, é o GLRT. Seja $N_d(x)$ a estatística de varrimento definida como $\sup_{0 < x < 1-d} n(x, d)$, sendo $n(x, d)$ o número máximo de pontos na janela de varrimento $]x, x+d]$ ($0 \leq x \leq 1-d$), d uma constante conhecida ($0 < d < 1$), e a e b parâmetros desconhecidos. $N_d(x)$ devolve um valor de x tal que o intervalo $]x, x+d]$ é o que possui maior número de eventos. Posto isto, o teste rejeita H_0 para valores elevados de N_d .

A estatística de teste é

$$\Lambda = \Lambda(X_1, \dots, X_N) = \frac{\sup_{\Theta_1} \prod_i f(X_i)}{\sup_{\Theta_0} \prod_i f(X_i)} \quad , \quad (1.4)$$

onde Θ_0 e Θ_1 são os subconjuntos do espaço paramétrico correspondentes, respetivamente, à hipótese nula e à hipótese alternativa. Sendo o denominador da equação (1.4) constante, podemos testar de forma equivalente utilizando

$$\begin{aligned} \Lambda &= \sup_{\Theta_1} \prod_i f(X_i) \\ &= \sup_{\Theta_1} a^n \left(\frac{1-ad}{1-d} \right)^{N-n} \quad , \end{aligned} \quad (1.5)$$

onde $\Theta_1 = \{(a, b, d) : 0 < d < 1, 1 < a \leq 1/d, 0 \leq b \leq 1 - d\}$ e $N = n(b, d)$.

Na expressão (1.5), quando d é uma constante conhecida, o supremo é atingido quando $a = n/N_d$, se $n \geq N_d$ e, caso contrário, se $a = 1$. Dada a restrição presente em Θ_1 ($1 < a \leq 1/d$), que nos limita à primeira situação, (1.5) pode ser simplificada para

$$\left(\frac{n}{N}\right)^n \left(\frac{N-n}{N}\right)^{N-n} \left(\frac{1}{d}\right)^n \left(\frac{1}{1-d}\right)^{N-n} \stackrel{def}{=} G(d, n) \quad . \quad (1.6)$$

Dado que $G(d, n)$ é monótona crescente em n , para um d fixo e $n \geq N_d$, o GLRT rejeita a hipótese nula quando n é mais elevado (Nagarawalla, 1996).

A Estatística de Varrimento com Janela Variável

Quando d deixa de ser um valor fixo e conhecido, já não se encontra uma variável aleatória com uma distribuição conhecida que permita a Λ ser uma função monótona dessa variável. Não obstante, para $n = 1, \dots, N - 1$ fixo, $G(d, n)$ é uma função convexa de d para $0 < d < 1$. O seu ínfimo é obtido quando $d = n/N$, e aproxima-se de ∞ à medida que $d \downarrow 0$. Aparentemente, rejeitar-se-ia sempre a hipótese nula, pois escolhendo um d pequeno e assegurando que n toma no mínimo o valor 1, Λ teria um valor elevado. Na prática, tal implica que cada ponto, por si, seria um aglomerado pelo simples facto de se encontrar no interior de um intervalo infinitesimalmente pequeno. Para permitir a deteção de *clusters* com uma duração de interesse prático pode-se limitar os valores que d toma como maiores que um determinado valor mínimo, ou limitar a análise para aglomerados com $n \geq n_0$ pontos, para um n_0 escolhido. A abordagem escolhida por Nagarwala, no qual se baseia a análise de *clusters* temporais do EUROCAT, passa por fixar a atenção em *clusters* com um tamanho mínimo predeterminado. Dado $n = n(b, d)$, o supremo da equação (1.5) pode ser encontrado para todo o $n \geq n_0$, ao invés de todos os valores possíveis de b (Nagarawalla, 1996).

Seja (X_1, \dots, X_N) uma amostra aleatória de N pontos com a função densidade de probabilidade descrita em (1.1), o GLRT para testar as hipóteses em (1.2), quando a , b , e d , são desconhecidos irá rejeitar H_0 para valores elevados da estatística de teste

$$\Lambda = \sup_{0 < d < n/N, n \geq n_0} \left(\frac{n}{N}\right)^n \left(\frac{N-n}{N}\right)^{N-n} \left(\frac{1}{d}\right)^n \left(\frac{1}{1-d}\right)^{N-n} \quad . \quad (1.7)$$

As possibilidades de aglomeração dos casos para os quais este teste foi feito correspondem a subintervalos de $(0, 1)$ de largura variável e desconhecida, no interior dos quais existe um risco elevado, ainda que uniforme. O teste considera somente *clusters* que contenham n_0 ou mais ocorrências (Nagarawalla, 1996).

A aplicação da estatística de varrimento de janela variável no EUROCAT

A estatística de varrimento λ , valor realizado de Λ apresentado em (1.7), é calculada para um conjunto de eventos organizados por data, ocorridos num determinado período temporal. Dado, para a maioria dos conjuntos de dados, o Λ descrito em (1.7) não possuir uma distribuição conhecida, e não ser possível determinar uma função g tal que $g(\Lambda)$ seja estritamente monótona com uma distribuição conhecida ou tabelada, é utilizada simulação de Monte Carlo para determinar a significância do λ obtido. Deste modo é encontrado um λ_{sig} , o 95º percentil do valor máximo de λ registado em cada simulação. Com um número elevado de ocorrências, devido à variabilidade da janela testada, são necessários numerosos testes, sendo que o método é ajustado automaticamente para estes testes múltiplos através da simulação de Monte Carlo. Não obstante, a janela variável permite uma maior sensibilidade para detetar mudanças na distribuição subjacente aos dados observados (Glaz & Zhang, 2004).

O algoritmo, adaptado daquele descrito por Nagarwalla (1996), encontra-se codificado em *Visual Basic for Applications* e disponível no EDMP, que permite utilização local em cada registo do EUROCAT. Quando possível, i.e., quando não existem demasiados valores omissos que impeçam este cálculo, o algoritmo utiliza a data estimada de conceção, e não a de nascimento, resultando na análise ser realizada para o período desde 1 de janeiro do primeiro ano até 31 de março do último ano considerado. De notar que o registo analisado deve ter uma população estável, sem uma alteração superior a 10% durante o intervalo temporal analisado (Teljeur et al., 2015).

O EDMP mostra os resultados com *clusters* agrupados, caso exista sobreposição temporal de 75% ou mais dos casos, e sejam ambos estatisticamente significativos. Após deteção de todos os aglomerados, aquele com maior valor de λ é colocado no primeiro grupo e os restantes aglomerados significativos são testados para sobreposição temporal com o mesmo. Os que tiverem 75% ou mais dos casos sobrepostos com o primeiro, serão incluídos no primeiro grupo. Em seguida, o processo é repetido para os *clusters* restantes, com a identificação do que possui o maior λ e agrupamento dos *clusters* que cumpram essa condição. Todo este processo é realizado de forma independente para as AC e grupos de AC considerados mais relevantes na vigilância epidemiológica. Exemplificando, os defeitos do SNC na sua totalidade (códigos CID-10: Q00, Q01, Q02, Q03, Q04, Q05, Q06, Q07), defeitos do tubo neural de um modo mais específico (códigos CID-10: Q00, Q01, Q05), ou ainda de forma ainda mais particular, espinha bífida (código CID-10: Q05) (Teljeur et al., 2015; EUROCAT, *website*).

Sendo objetivo do RENAC o estudo e análise da situação nacional no que respeita às AC, e como membro do EUROCAT, ao enviar e tratar os dados de AC presentes em nascimentos ocorridos em grávidas que residem em localidades situadas a sul do Tejo para análise central, utiliza o *software* EDMP. No entanto, este é insuficiente para os objetivos do RENAC e permite apenas a análise temporal de *clusters*.

1.6 A estatística de varrimento espacial e espaço-temporal

A estatística de varrimento, já referida, é utilizada principalmente para detetar um excesso delimitado de eventos ou ocorrências, e testar se esse excesso ocorreu por acaso. As propriedades da estatística de varrimento incluem: a área a ser analisada, a distribuição de probabilidade que reflete as ocorrências sob a hipótese nula, e a forma e tamanho da janela de varrimento utilizada. Dependendo da aplicação desejada podem ser utilizados diferentes modelos estatísticos e, com base em cada modelo, a estatística de teste pode ser avaliada de forma direta ou com recurso à simulação de Monte Carlo. Nesta última situação, são gerados conjuntos aleatórios de dados sob a hipótese nula, e comparados de imediato com o conjunto de dados real para realizar então o teste de hipóteses. Tal dirá quão provável é o conjunto real de dados verificar-se se a hipótese nula for verdadeira (Kulldorff, 1999).

Um modelo geral para a estatística de varrimento espacial e espaço-temporal

Seja A o espaço no qual os eventos podem ocorrer, tratando-se de uma porção do espaço euclidiano em que as diferentes dimensões podem representar quer o espaço, quer o tempo. Em A , defina-se μ , que será a intensidade subjacente à geração de eventos sob a hipótese nula. Para um processo de Poisson homogéneo, tem-se $\mu(x) = \lambda$ para todo o $x \in A$. Tomando X como um processo pontual no espaço, e B um subconjunto de A , tal que $B \subset A$, $X(B)$, será então um número aleatório de eventos em B . Partindo daqui, obtêm-se dois modelos de probabilidade distintos, um baseado em contagens de Bernoulli e outro no processo de Poisson (Kulldorff, 1997; Kulldorff, 1999).

Para o modelo de Bernoulli serão consideradas apenas medidas discretas, μ , de tal forma que $\mu(B)$ é um número inteiro para todos os subconjuntos $B \subset A$. Cada unidade de medida corresponde a uma entidade ou indivíduo com um de dois estados possíveis, a título de exemplo, ter uma determinada doença ou não. Um destes estados é o evento de interesse, e a localização dos indivíduos considerados no espaço é o processo pontual. Sob a hipótese nula, tem-se que o número de eventos numa qualquer área possui uma distribuição binomial, tal que

$$X(B) \stackrel{H_0}{\sim} \text{Bin}(\mu(B), p) \quad , \quad (1.8)$$

para um valor de p e todos os subconjuntos $B \subset A$ (Kulldorff, 1997; Kulldorff, 1999).

Passando ao modelo de Poisson, os eventos são gerados por processo de Poisson homogéneo ou não-homogéneo. Sob a hipótese nula

$$X(B) \stackrel{H_0}{\sim} \text{Poisson}(p\mu(B)) \quad , \quad (1.9)$$

para um valor de p e todos os subconjuntos $B \subset A$. A medida μ pode ser definida de forma contínua, de tal modo que os eventos podem ocorrer em qualquer lugar, ou de uma forma discretizada, fazendo com que os eventos só possam ocorrer em localizações específicas, ou ainda uma combinação de ambos os casos. A definição discreta pode ser particularmente útil ao trabalhar com contagens individuais ou dados agregados (Kulldorff, 1997; Kulldorff, 1999).

A janela de uma estatística de varrimento pode ser pensada como um intervalo, uma área, ou um volume fixo no seu tamanho e na sua forma, que se move ao longo da área de estudo. Ao percorrer a área de estudo define a coleção \mathcal{W} de regiões $W \subset A$. Generalizando, permitem-se janelas de diferentes formas e tamanhos ao definir a janela como uma coleção \mathcal{W} de região $W \subset A$ de qualquer forma e tamanho. O que a torna numa janela de varrimento é o facto de diferentes regiões se sobreporem, e, no seu conjunto, cobrirem A . (Kulldorff & Nagarwalla, 1995; Kulldorff, 1997; Kulldorff, 1999)

Com o número total de eventos observados, $X(A)$, enquanto condição, a definição da estatística de varrimento será então a razão da máxima verosimilhança sobre todas as regiões possíveis:

$$S_{\mathcal{W}} = \frac{\max_{W \in \mathcal{W}} L(W)}{L_0} = \max_{W \in \mathcal{W}} \frac{L(W)}{L_0} \quad , \quad (1.10)$$

onde $L(W)$ é a função de verosimilhança para a região W , que expressa o quão provável é os dados observados terem uma proporção diferente no interior face ao exterior dessa mesma região, e L_0 é a função de verosimilhança sob a hipótese nula (Kulldorff, 1999).

Denote-se $X(A/W) = X(A) - X(W)$ e $\mu(A/W) = \mu(A) - \mu(W)$. Desenvolvendo a equação (1.10) para o modelo de Bernoulli obtém-se

$$\frac{L(W)}{L_0} = \frac{\left(\frac{X(W)}{\mu(W)}\right)^{X(W)} \left(1 - \frac{X(W)}{\mu(W)}\right)^{\mu(W)-X(W)} \left(\frac{X(A/W)}{\mu(A/W)}\right)^{X(A/W)} \left(1 - \frac{X(A/W)}{\mu(A/W)}\right)^{\mu(A/W)-X(A/W)}}{\left(\frac{X(A)}{\mu(A)}\right)^{X(A)} \left(1 - \frac{X(A)}{\mu(A)}\right)^{\mu(A)-X(A)}} \quad , \quad (1.11)$$

se $X(W)/\mu(W) > X(A/W)/\mu(A/W)$. Caso contrário, $L(W) = 1$.

Considerando o modelo de Poisson tem-se

$$\frac{L(W)}{L_0} = \frac{\left(\frac{X(W)}{\mu(W)}\right)^{X(W)} \left(\frac{X(A/W)}{\mu(A/W)}\right)^{X(A/W)}}{\left(\frac{X(A)}{\mu(A)}\right)^{X(A)}} \quad , \quad (1.12)$$

se $X(W)/\mu(W) > X(A/W)/\mu(A/W)$. Caso contrário $L(W) = 1$. A expressão $X(W)/\mu(W) >$

$X(A/W)/\mu(A/W)$ indica que no interior da janela existem mais casos que os esperados, comparando com o exterior da janela. Para encontrar áreas com menores proporções ou taxas, a expressão seria $X(W)/\mu(W) < X(A/W)/\mu(A/W)$. Quando o tamanho da janela é fixo no que toca ao número de casos esperados, ou seja, $\mu(W) = \mu(W')$ para todos os $W, W' \in \mathcal{W}$, a estatística de varrimento torna-se:

$$S'_{\mathcal{W}} = \max_{W \in \mathcal{W}} X(W), \quad (1.13)$$

i.e., o número máximo de eventos no interior das janelas sobre todas as localizações possíveis. Apesar de $S'(\mathcal{W}) \neq S(\mathcal{W})$, para duas quaisquer concretizações, denominadas, sem perda de generalidade, por w_1 e w_2 , verificar-se-à $S'_{\mathcal{W}}(w_1) > S'_{\mathcal{W}}(w_2)$ se e só se $S_{\mathcal{W}}(w_1) > S_{\mathcal{W}}(w_2)$. De outra forma, tal quer dizer que se a janela for fixa, um teste de hipóteses baseado em $S'_{\mathcal{W}}$ é idêntico a um baseado em $S_{\mathcal{W}}$ (Kulldorff, 1999).

Partindo do acima descrito, basta adicionar o tempo na escala discreta mais adequada para o objeto de estudo em apreço (dia, mês, ou ano), como uma terceira dimensão, para estudar *clusters* espácio-temporais, tendo as restantes duas dimensões espaciais, no plano. Tal é feito através da junção do círculo espacial (duas dimensões) com o intervalo temporal (uma terceira dimensão) que resulta numa janela cilíndrica que pode ser feita variar na sua área, ou seja, na base do cilindro, e no período temporal estudado, a altura do cilindro (Kulldorff, 1999).

1.7 Objetivos

1.7.1 Objetivo principal

A finalidade deste projeto compreende a criação de uma ferramenta computacional que permita uma análise célere de *clusters* espaço-temporais, para utilização no âmbito da vigilância das AC em Portugal. Deve ser flexível, a fim de permitir a análise de AC específicas, ou grupos de AC, a análise de períodos temporais, e a análise de regiões de Portugal, ambos escolhidos de acordo com os critérios do utilizador.

1.7.2 Objetivos secundários

Além da necessidade de investigar uma metodologia que se adeque à análise estatística de *clusters* espaço-temporais de casos de AC em Portugal, é também necessário encontrar uma forma automatizada de trabalhar os dados existentes para permitir o uso expedito e simplificado da mesma.

Para ilustrar a utilidade e aplicabilidade da ferramenta, também se pretende fazer uso da mesma numa análise de AC e grupos de AC relevantes para a vigilância epidemiológica, tendo como referência as escolhidas pelo EUROCAT.

Capítulo 2

Materiais e métodos

2.1 Dados utilizados

2.1.1 Dados das anomalias congénitas

Os dados do RENAC analisados neste trabalho compreendem todos os casos de AC que foram notificados entre o início do ano 2000 e o final do ano 2013. A base de dados foi extraída, através da plataforma RIOS, no formato Folha de Cálculo do *Microsoft Excel*, num total de sete folhas, cada uma compreendendo informações distintas: a primeira folha compila diversas informações sobre o indivíduo afetado, a gestação, comportamentos maternos e o momento de diagnóstico da(s) anomalia(s); a segunda folha, compreende informações sobre a presença de doença materna aguda no 1º trimestre da gravidez; a terceira folha inclui variáveis sobre a presença de doença crónica materna; a quarta folha fornece informação sobre a(s) anomalia(s) que o indivíduo possui; a quinta e sexta folhas identificam AC que, respetivamente, o pai e a mãe do indivíduo tenham; a sétima e última folha indica medicamentos que a mãe tenha tomado durante o primeiro trimestre da gravidez. No Anexo 2 encontram-se descritas em pormenor todas as variáveis presentes na base de dados do RENAC.

Codificação das anomalias congénitas

O RENAC codifica as AC de acordo com a 10ª Revisão da Classificação Internacional de Doenças (CID-10), utilizando para tal o Capítulo XVII da mesma, ao qual acrescentou algumas AC que estão codificadas fora deste capítulo de acordo com os critérios de codificação e qualidade dos dados previstos no EUROCAT (Departamento de Epidemiologia, 2011).

Nas tabelas seguintes apresentam-se respetivamente: a distribuição e codificação das AC por grandes grupos assim como aquelas que foram incluídas, embora pertencentes a outros capítulos (WHO,

2015).

Tabela 2.1: Codificação das AC por grandes grupos, de acordo com o capítulo XVII da CID-10, utilizada pelo RENAC entre 2000 e 2013.

Código	Descrição
Q00-Q07	AC do Sistema Nervoso Central
Q10-Q18	AC do Olho, Ouvido, Face e Pescoço
Q20-Q28	AC do Aparelho Circulatório
Q30-Q34	AC do Aparelho Respiratório
Q35-Q37	Fenda Labial e Fenda Palatina
Q38-Q45	Outras AC do Aparelho Digestivo
Q50-Q56	AC do Aparelho Genital
Q60-Q64	AC do Aparelho Urinário
Q65-Q79	AC e Deformações do Sistema Músculo-Esquelético
Q80-Q89	Outras AC
Q90-Q99	AC Cromossómicas não classificadas noutra local

Tabela 2.2: Codificação das AC não incluídas no capítulo XVII da CID-10, utilizada pelo RENAC entre 2000 e 2013, e o grupo do mesmo capítulo em que as mesmas foram inseridas.

Código	Descrição	Inseridas em
C41.4	Teratoma sacrococccígeo	Q80-Q89
C49.0	Teratoma cervical	Q80-Q89
C49.4	Teratoma abdominal	Q80-Q89
D18.00	Hemangioma em local não especificado	Q80-Q89
D18.01	Hemangioma da pele e tecido subcutâneo	Q80-Q89
D18.10	Higroma quístico	Q10-Q18
D82.1	Síndrome de Di George	Q80-Q89
K07.0	Micrognatia/Retrognatia	Q10-Q18
K07.1	Prognatismo	Q10-Q18
K40	Hérnia inguinal	Q65-Q79
K42	Hérnia umbilical	Q65-Q79
P75	Íleo meconial	Q38-Q45
P83.2	Hidropsis fetal	Q20-Q28
P96.3	Sutura sagital aberta	Q65-Q79

2.1.2 Dados populacionais

Neste estudo, foram utilizados dados populacionais fornecidas pelo Instituto Nacional de Estatística (INE), em https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE, nomeadamente, os valores totais de nascimentos, i.e., o número de fetos mortos somado ao número de nados-vivos,

para cada concelho de Portugal, por mês. Estes dados foram recolhidos num ficheiro no formato Folha de Cálculo do *Microsoft Excel* para cada ano considerado, de 2000 a 2013. De seguida se apresenta um exemplo destes ficheiros.

	A	B	C	D	E	F	G	H
1								
2	Quadro extraído em 18 de Maio de 2017 (15:45:23)							
3	http://www.ine.pt							
4								
5								
6	Período de referência dos dados (1)	Lista de meses	Local de residência da mãe (NUTS - 2002) (2)		Óbitos fetais (N.º) por Local de residência da mãe (NUTS - 2002) e Mês (óbito); Anual		Nados-vivos (N.º) por Local de residência da mãe (NUTS - 2002), Sexo e Mês de nascimento; Anual (3)	
7								Sexo
8								HM
9								N.º
10	2000	Janeiro	Portugal	PT		45		9825
11			Continente	1		43		9292
12			Arcos de V	1111601	✓	0		12
13			Caminha	1111602	✓	0		12
14			Melgaço	1111603	✓	0		5
15			Monção	1111604	✓	0		12
16			Paredes de	1111605	✓	0		10
17			Ponte da B	1111606	✓	0		13
18			Ponte de Li	1111607	✓	0		37
19			Valença	1111608	✓	0		11
20			Viana do C	1111609	✓	0		75

Figura 2.1: Exemplo do ficheiro com os valores totais de nascimentos para cada concelho de Portugal, por mês, referentes ao ano 2000.

2.1.3 Coordenadas dos concelhos de Portugal

A fim de incluir a componente espacial pretendida na análise de *clusters* deste trabalho, foram utilizadas as coordenadas geográficas de latitude e de longitude dos centroides dos concelhos de Portugal. O cálculo dos centroides dos concelhos foi feito com recurso ao *software* ArcGIS, um sistema de informação geográfica, em inglês *geographical information system* (GIS). Estes dados foram colocados em cinco ficheiros de texto com os valores separados por vírgulas, um para cada região geográfica de interesse, a saber: Portugal continental, arquipélago da Madeira, grupo ocidental do arquipélago dos Açores, grupo central do arquipélago dos Açores, e grupo oriental do arquipélago dos Açores. Todos estes ficheiros possuem as mesmas variáveis, codificadas do mesmo modo, enumerando: o identificador do campo (um contador que inicia no 0 e acresce uma unidade por linha, denominado 'FID'), o nome do concelho, o código DDCC do concelho (identificador numérico do

concelho em que os dois primeiros dígitos identificam o distrito e os dois seguintes o concelho), a longitude, e a latitude do seu centroide. A figura que se segue é um excerto do ficheiro com os dados geográficos para Portugal continental.

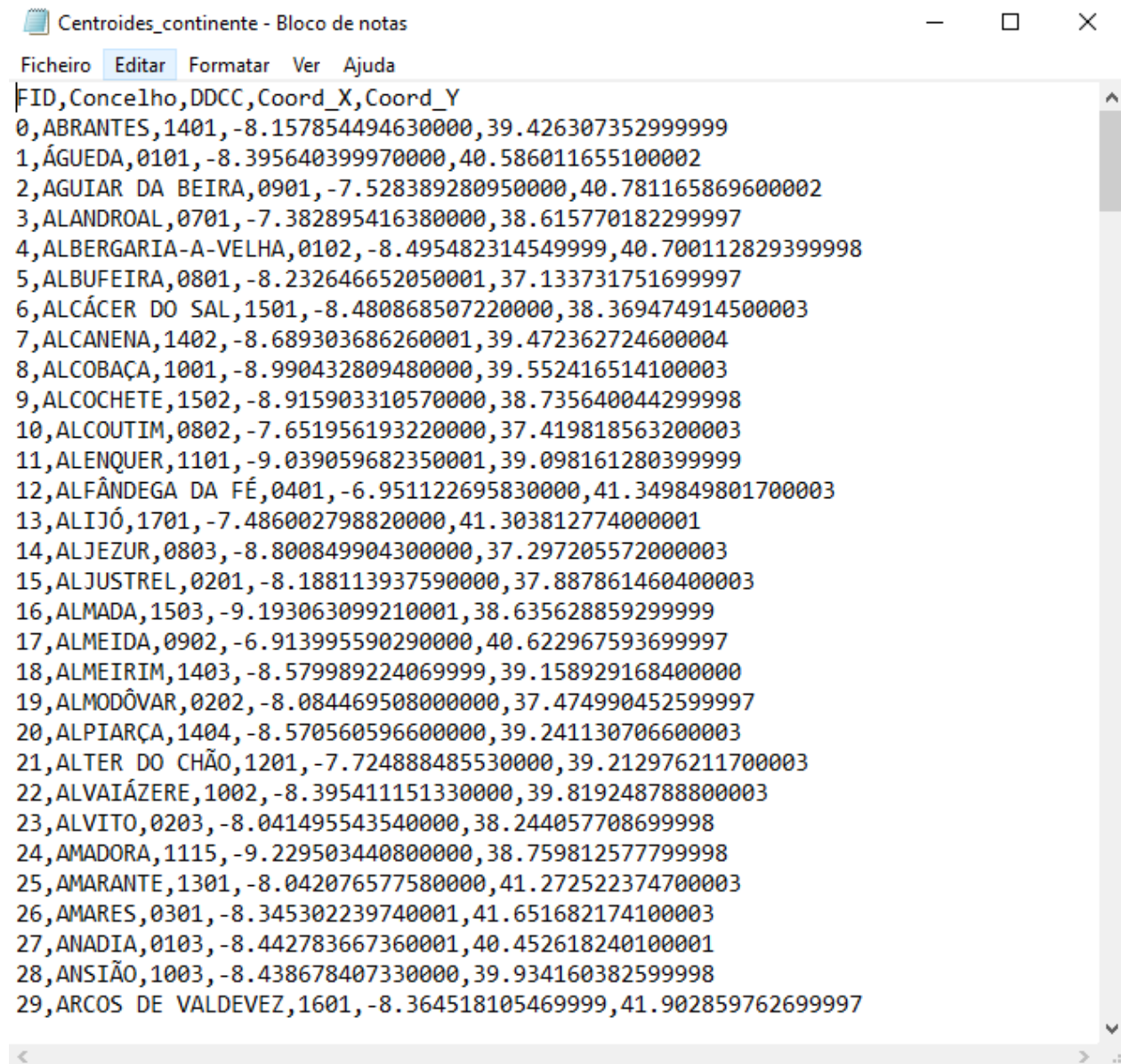


Figura 2.2: Excerto do ficheiro com os dados geográficos para os municípios de Portugal continental.

A divisão do território nacional nas cinco regiões apresentadas deve-se à natureza da análise e da metodologia usada, exposta mais à frente. De um modo geral, os pontos populacionais considerados, neste caso os centroides dos concelhos, são usados para criar círculos de dimensão cada vez maior, cada um contendo mais um ponto populacional. Dada a distância entre alguns destes pontos de diferentes regiões e a heterogeneidade nas dimensões geográficas e populacionais, será mais correto uma pesquisa separada para cada uma das cinco regiões.

2.2 Metodologia

A análise de *clusters* espaço-temporais de AC foi efetuada com recurso a *software* de livre acesso, nomeadamente: **R** (R Core Team, 2016) e SaTScanTM (SaTScan, *website*), bem como a um *package* para o **R**, o *rsatscan* (Kleinman, 2015), que permite a utilização do SaTScanTM através da interface do **R**. Para correr o algoritmo desejado do SaTScanTM é necessário introduzir de um modo específico as tabelas que contêm os dados mencionados no início deste capítulo. Com base nos dados acima descritos, foi criado um algoritmo em **R** que os transforma nessas mesmas tabelas e cujo *script* se encontra integralmente no Anexo 3.

As três matrizes de dados necessárias à utilização do SaTScanTM resultam da filtragem de, pelo menos, quatro ficheiros, que correspondem: à primeira folha da base de dados do RENAC, onde se encontra a informação espaço-temporal dos casos, à quarta folha da mesma base de dados, que contém os dados sobre as AC dos casos, ao ficheiro com os dados espaço-temporais dos nascimentos ocorridos em Portugal por ano, fornecido pelo INE (caso seja considerado um período temporal maior que um ano será preciso um ficheiro para cada ano), e ao ficheiro com os dados das coordenadas geográficas dos centroides dos concelhos de Portugal. A forma como estes ficheiros são transformados nas matrizes de dados aceites pelo SaTScanTM será explicada mais adiante.

As três tabelas de dados têm que possuir as variáveis referidas em seguida, na ordem apresentada, à qual se junta um exemplo do cabeçalho das mesmas, e são:

- casos, com as variáveis: identificação da localização espacial, número de casos para uma dada localização e mês e ano do nascimento;

	Concelho	Casos	Data
1	abranter	0	01-2013
2	abranter	0	02-2013
3	abranter	0	03-2013
4	abranter	0	04-2013
5	abranter	0	05-2013
6	abranter	0	06-2013
7	abranter	0	07-2013
8	abranter	0	08-2013
9	abranter	0	09-2013
10	abranter	0	10-2013
11	abranter	0	11-2013
12	abranter	0	12-2013
13	águeda	0	01-2013
14	águeda	0	02-2013

Figura 2.3: Exemplo da matriz de dados dos casos como vista na interface do **R**.

- controlos, com as variáveis: localização espacial, número de controlos (“CTL”) para uma dada localização e mês de nascimento, e o mês e ano do nascimento;

	Concelho	CTL	Data
1	abrantas	21	01-2013
2	abrantas	16	02-2013
3	abrantas	12	03-2013
4	abrantas	14	04-2013
5	abrantas	15	05-2013
6	abrantas	14	06-2013
7	abrantas	21	07-2013
8	abrantas	23	08-2013
9	abrantas	19	09-2013
10	abrantas	22	10-2013
11	abrantas	9	11-2013
12	abrantas	28	12-2013
13	águeda	27	01-2013
14	águeda	29	02-2013

Figura 2.4: Exemplo da matriz de dados dos controlos como vista na interface do R.

- coordenadas geográficas, com as variáveis: identificação do concelho, e a latitude e longitude do mesmo.

	Concelho	Latitude	Longitude
1	abrantas	39.426307353	-8.15785449463
2	águeda	40.5860116551	-8.39564039997
3	aguiar_da_beira	40.7811658696	-7.52838928095
4	alandroal	38.6157701823	-7.38289541638
5	albergaria-a-velha	40.7001128294	-8.49548231455
6	albufeira	37.1337317517	-8.23264665205
7	alcácer_do_sal	38.3694749145	-8.48086850722
8	alcanena	39.4723627246	-8.68930368626
9	alcobaça	39.5524165141	-8.99043280948
10	alcochete	38.7356400443	-8.91590331057
11	alcoutim	37.4198185632	-7.65195619322
12	alenquer	39.0981612804	-9.03905968235

Figura 2.5: Exemplo da matriz de dados com as informações geográficas como vista na interface do R.

As duas primeiras tabelas são extensas, havendo, para cada ano de análise na região de Portugal continental, 3336 entradas (o produto do número de concelhos em Portugal continental, 278, pelos 12 meses de cada ano). A terceira tabela possui, por conseguinte, para Portugal continental, 278 entradas.

Para o problema específico desta análise, a identificação da localização espacial refere-se ao nome do concelho de residência da mãe do indivíduo com AC, dado permitir uma análise mais fina do que a do distrito, e a informação sobre a freguesia se encontrar incompleta em alguns registos (aproximadamente 30% de valores omissos de 2000 a 2013). A somar-se a esta situação, a descrição de um *cluster* espaço-temporal, por freguesia, de uma AC com efeitos físicos ou estéticos muito evidentes pode levar à identificação dos indivíduos afetados, negando o propósito do anonimato dos indivíduos na base de dados do RENAC.

A informação temporal para a análise efetuada é dada pelo ano e mês de nascimento, e o número de controlos é obtido subtraindo ao número de nascimentos obtidos nas tabelas fornecidas pelo INE para os anos em estudo, o total de casos de AC por data de nascimento dos casos e concelho de residência da mãe.

A estatística de varrimento espacial e espaço-temporal na qual o SaTScanTM se baseia já foi apresentada na secção 1.6, sendo de seguida explicada de um modo mais pormenorizado e aplicado de forma prática. Também será exposto, em conjunto, o modelo de Bernoulli, como parte da estatística de varrimento espacial e espaço-temporal, modelo este escolhido para a análise dos dados, sendo o evento de interesse a ocorrência de um caso com AC. A escolha do modelo de Bernoulli, face por exemplo ao modelo de Poisson, para a análise incluída na ferramenta descrita neste trabalho, deve-se à natureza dos casos de AC se adaptarem melhor a este; para utilizar o modelo de Poisson, seria necessário ter informação sobre a população em risco ao longo do tempo, i.e., sobre as mulheres grávidas, desde o momento da conceção ao momento do parto. Assim, à semelhança daquilo que já foi feito em estudos semelhantes sobre AC, e outros eventos que podem ocorrer durante a gravidez (Ozdenrol, Williams, Kang & Magsumbol, 2005; Viel, Floret & Mauny, 2005; Kuehl & Loffredo, 2006.), será utilizado o modelo de Bernoulli.

De notar que foram considerados, para esta análise, alguns métodos de *clustering* contemplados na secção 1.4, tendo estes sido excluídos devido ao facto de a sua principal utilização ser a redução da dimensionalidade de base de dados em *Data Mining*, ao agrupar os elementos das mesmas em classes, e ao facto de um dos principais propósitos do SaTScanTM ser o de realizar análises de eventos relevantes em Saúde Pública, tendo já sido amplamente usado e testado, desde a sua criação, em estudos sobre, por exemplo, doença oncológica (Henry, Niu & Boscoe, 2009), doença respiratória (Park, Yoon, Na & Song, 2015) e doença neurológica (Banta, Addison & Beeson, 2015), além dos já mencionados no parágrafo supra.

Recordando: X é um processo pontual no espaço; A é o espaço onde eventos podem ocorrer; $X(B)$ é o número aleatório de pontos no subconjunto $B \subset A$; \mathcal{W} é uma coleção de regiões $W \subset A$; W é utilizado de forma permutável para definir tanto um subconjunto de A , como os parâmetros

referentes à região em questão; $\mu(B)$ é a unidade de medida e, para o modelo de Bernoulli, é um número inteiro para todo o $B \subset A$, que corresponde a uma entidade ou indivíduo e indica qual o seu estado, e.g., se possui a doença ou não; um indivíduo ou entidade num dos dois estados possíveis é considerado um ponto, e a sua localização espacial ou espaço-temporal é o processo pontual (Kulldorff, 1997).

De acordo com o modelo, existe uma região $W \subset A$, tal que cada indivíduo ou entidade no seu interior tem a probabilidade p de ser um ponto, enquanto a probabilidade fora da região é q . O acontecimento que se define como “um indivíduo ser um ponto de interesse” é independente dos restantes indivíduos (Kulldorff, 1997).

A hipótese nula é

$$H_0 : p = q \quad , \quad (2.1)$$

ou seja, a probabilidade de um indivíduo no interior da região $W \subset A$ se encontrar no estado de interesse é igual à probabilidade de um indivíduo no exterior se encontrar nesse mesmo estado. Logo, a hipótese alternativa é

$$H_1 : p > q, W \in \mathcal{W} \quad , \quad (2.2)$$

isto é, a probabilidade de um indivíduo no interior da região $W \subset A$ se encontrar no estado de interesse é maior que no seu exterior.

Logo, sob H_0 ,

$$X(B) \stackrel{H_0}{\sim} \text{Bin}(\mu(B), p) \quad , \quad (2.3)$$

para todos os conjuntos B . Sob H_1 ,

$$X(B) \stackrel{H_1}{\sim} \text{Bin}(\mu(B), p) \quad , \quad (2.4)$$

para todos os conjuntos $B \subset W$, e

$$X(B) \stackrel{H_1}{\sim} \text{Bin}(\mu(B), q) \quad , \quad (2.5)$$

para todos os conjuntos $B \subset \overline{W}$ (Kulldorff, 1997).

O teste da razão de verosimilhanças

A função de verosimilhança para o modelo de Bernoulli é

$$L(W, p, q) = p^{x_W} (1-p)^{\mu(W)-x_W} q^{x_A-x_W} (1-q)^{[\mu(A)-\mu(W)]-(x_A-x_W)} \quad , \quad (2.6)$$

denotando x_A e x_W , respetivamente, o número total de pontos observados e o número total de pontos observados no interior da região W .

Para detetar a região que possui maior probabilidade de ser um *cluster*, é necessário encontrar a região \hat{W} que maximiza a função de verosimilhança. Este \hat{W} é, assim, o estimador de máxima verosimilhança do parâmetro W . Maximizando a função de verosimilhança condicionada a W obtém-se

$$\begin{aligned} L(W) \stackrel{def}{=} \sup_{p>q} L(W, p, q) &= \left(\frac{x_W}{\mu(W)} \right)^{x_W} \left(1 - \frac{x_W}{\mu(W)} \right)^{\mu(W)-x_W} \\ &\times \left(\frac{x_A - x_W}{\mu(A) - \mu(W)} \right)^{x_A - x_W} \left(1 - \frac{x_A - x_W}{\mu(A) - \mu(W)} \right)^{[\mu(A) - \mu(W)] - (x_A - x_W)} \quad , \end{aligned} \quad (2.7)$$

quando $\frac{x_W}{\mu(W)} > \frac{x_A - x_W}{\mu(A) - \mu(W)}$. Caso contrário,

$$L(W) = \left(\frac{x_A}{\mu(A)} \right)^{x_A} \left(\frac{\mu(A) - x_A}{\mu(A)} \right)^{\mu(A) - x_A} \quad . \quad (2.8)$$

Denote-se

$$L_0 \stackrel{def}{=} \sup_{p=q} L(W, p, q) = \left(\frac{x_A}{\mu(A)} \right)^{x_A} \left(\frac{\mu(A) - x_A}{\mu(A)} \right)^{\mu(A) - x_A} \quad . \quad (2.9)$$

Posto isto, a razão de verosimilhanças, λ , pode ser escrita como

$$\lambda = \frac{\sup_{Z \in \mathcal{Z}, p>q} L(W, p, q)}{\sup_{p=q} L(W, p, q)} = \frac{L(\hat{W})}{L_0} \quad . \quad (2.10)$$

Note-se que o denominador não está dependente da distribuição espacial dos pontos, mas apenas do seu número, x_A . A razão de verosimilhanças descrita é a estatística de teste utilizada (Kulldorff, 1997).

Inferência estatística sobre os *clusters* detetados

Além da detecção dos *clusters*, é ainda necessário efetuar inferência estatística sobre os mesmos. Deste modo, além de se localizar a região que obteve o resultado de *cluster*, é também necessário rejeitar a hipótese nula, que afirma a inexistência de *clusters*, para um determinado nível de significância.

Seja $n = \{n_i, i = 1, \dots, x_A\}$ o conjunto de coordenadas dos x_A pontos presentes nos dados, onde \hat{W} é o *cluster* com maior probabilidade associada, e $n' = \{n'_i, i = 1, \dots, x_A\}$ uma qualquer configuração alternativa dos mesmos x_A pontos. A construção da estatística de varrimento envolve um teorema que estabelece que se a hipótese nula é rejeitada para n , é rejeitada também para n' , se $n'_i = n_i$ para todo o $n_i \in \hat{W}$. Quer isto dizer que, desde que os pontos no interior da região que constitui o *cluster* mais provável estejam localizados efetivamente no seu interior, os restantes, fora da região, podem encontrar-se dispostos aleatoriamente. Para provar este teorema, sejam $\lambda(n)$ e $\lambda(n')$ os valores da estatística de teste para dois conjuntos diferentes de dados. Dado estes conjuntos terem o mesmo número de pontos, a distribuição de $X(B)$ sob H_0 é a mesma, e tal é suficiente para suportar a conclusão que $\lambda(n') \geq \lambda(n)$. Para o modelo de Bernoulli,

$$\lambda(n) = \frac{L(\hat{W}|n)}{L_0} \leq \frac{L(\hat{W}|n')}{L_0} \leq \frac{\sup_W L(W|n')}{L_0} = \lambda(n') \quad . \quad (2.11)$$

A primeira inequação é verdadeira dado n' ter tantos ou mais pontos no interior de W como n . Deste modo é possível aliar a detecção e localização de *clusters* à inferência estatística sobre as mesmas (Kulldorff, 1997).

Para obter o valor da estatística de teste é necessário calcular a razão de verosimilhanças à medida que esta é maximizada para uma coleção de regiões sob a hipótese alternativa. De notar que o facto de o número observado de pontos ser sempre finito e, para um número fixo de pontos, o valor da verosimilhança diminuir à medida que a janela de varrimento aumenta nas suas dimensões, permite a resolução concreta destes cálculos, ao invés do que acontece no seu abstrato em que o número de regiões poderia ser infinito. Sem perda de generalidade, considerando uma janela circular com centro num determinado ponto ou foco, só é necessário calcular a razão de verosimilhanças de cada vez que um ponto passa a encontrar-se no interior da janela, à medida que esta aumenta de dimensão, repetindo o processo para todos os pontos ou focos possíveis.

Como não é expectável encontrar uma distribuição subjacente com função de distribuição de probabilidade conhecida, calculado o valor da estatística de teste, recorre-se ao método de simulação Monte Carlo. Como a medida subjacente, μ , é conhecida, podem ser obtidas réplicas do conjunto de dados geradas sob H_0 , condicionadas a $n(A)$, o número total de pontos ou ocorrências. Realizadas 9999 réplicas, pode ser inferido que o *cluster* detetado é estatisticamente significativo ao nível de significância de $\alpha = 0,05$, se o valor da estatística de teste do mesmo se encontrar entre os 500 maiores valores da estatística de teste calculados a partir das réplicas (Kulldorff, 1997).

Com esta metodologia, além do *cluster* principal, são também detetados *clusters* secundários, com a respetiva inferência estatística associada, podendo estes *clusters* secundários ser de dois tipos, caso tenham sobreposição, ou não, ao *cluster* principal. No entanto, o SaTScanTM realiza a inferência estatística sobre os *clusters* secundários comparando o seu valor da estatística de teste com os valores da estatística de teste do *cluster* principal detetado nas réplicas do conjunto de dados, sendo por isso conservador na sua validação estatística.

O algoritmo do SaTScanTM

Apresenta-se de seguida o algoritmo presente no SaTScanTM para a deteção e validação estatística de *clusters* espaciais:

1. no conjunto de dados, os pontos presentes são dispostos numa grelha;
2. é escolhido um ponto;
3. para esse ponto são calculadas as distâncias aos restantes pontos;
4. os valores obtidos destas distâncias são ordenados de forma crescente e guardados numa matriz;
5. os passos 1, 2, 3 e 4 são repetidos para todos os pontos na grelha;
6. é escolhido, de novo, um ponto na grelha;
7. tendo o centro no ponto escolhido no passo 6, são criados círculos com um raio cada vez maior;
8. para cada ponto populacional que passa a estar incluído no círculo, é atualizado o número de eventos e a medida $\mu(W)$ no interior da área W ;
9. em todos os pontos, são repetidos os passos 6, 7 e 8;
10. a cada repetição é reportado o maior valor de verosimilhança com base em todos os pares $(n, \mu(W))$, sendo n o número de eventos no interior da região W ;
11. são repetidos os passos do 5 ao 10 para cada replicação Monte Carlo.

Na sua aplicação espaço-temporal, o SaTScanTM adiciona o intervalo temporal como uma terceira dimensão, formando-se uma janela cilíndrica, já mencionada na secção (1.6). Desta forma, com base no algoritmo apresentado supra, a grelha mencionada passa a ser tridimensional, e, para cada círculo criado no passo 7, é tida também em conta esta terceira dimensão, sendo o resultado uma janela de varrimento espaço-temporal. Os restantes passos e cálculos efetuados mantêm-se inalterados (Kulldorff, 1999).

2.3 Funcionamento da ferramenta

Na ferramenta encontra-se um conjunto de parâmetros introduzidos pelo utilizador que determinam vários aspetos importantes da análise a realizar. É definido pelo utilizador:

- o período temporal a considerar para a análise, que pode ir desde um ano até a todos os anos para os quais haja informação;
- a região de Portugal a analisar, que poderá ser Portugal continental, o arquipélago da Madeira, os grupos ocidental, central ou oriental do arquipélago dos Açores;
- os casos de AC a estudar, podendo ser analisados todos os casos notificados independentemente da AC observada, apenas casos portadores de uma qualquer AC de um dos grandes grupos mencionados na tabela 2.1, ou casos com AC de um subgrupo mais específico, e ainda casos com uma AC em particular;
- o limite máximo da componente temporal da janela de varrimento, em número de meses ou percentagem do período temporal analisado;
- o limite máximo da componente espacial da janela de varrimento, em percentagem de cobertura de nascimentos.

De acordo com os parâmetros introduzidos pelo utilizador, as matrizes de dados descritas no capítulo 2.2 são modificadas na sua forma, e encontram-se prontas a ser utilizadas para efetuar a análise de *clusters* espaço-temporais. Para tal, é utilizado um *package* do **R**, o *rsatscan*, que permite a utilização do SaTScanTM a partir da interface do **R**. No Anexo 4 encontra-se um manual de instruções que explica em pormenor como devem ser inseridos os parâmetros e o formato em que se devem encontrar as tabelas originais dos dados para poder utilizar esta ferramenta.

2.3.1 A inserção dos dados

Na pasta onde se encontram os ficheiros necessários para utilizar a ferramenta, indicada pelo utilizador ao realizar uma análise (e.g. 'C:\\Users\\afons\\Desktop\\Dados'), devem estar os seguintes ficheiros:

- a primeira folha do ficheiro com a base de dados do RENAC, com a coluna 'Comentario' eliminada, guardada em formato de valores delimitados por tabulação e com o nome 'cas1';
- a quarta folha do mesmo ficheiro, guardada também no formato de valores delimitados por tabulação e com o nome 'cas2';

- os ficheiros, um para cada ano, com os dados fornecidos pelo INE, com as informações mencionadas na secção 2.1.2 (ver figura 2.1), sem quaisquer alterações, guardados no formato *comma separated values* (CSV) com o nome ‘pop####’, sendo os # substituídos pelo ano a que se refere cada ficheiro (e.g. para o ano de 2000 o nome do ficheiro deve ser ‘pop2000’), referentes aos anos para os quais se pretende realizar a análise;
- os ficheiros de texto, com os valores separados por vírgulas, contendo as coordenadas geográficas dos concelhos da região para a qual se pretende realizar a análise (ver figura 2.2).

Mais informações sobre o formato em que estes ficheiros se devem encontrar, o nome que lhes está associado, as variáveis que devem conter e a forma como deve ser escrito o caminho da pasta que os contém, encontram-se no Anexo 4.

Além dos parâmetros já mencionados nesta secção, devem ainda ser indicados o caminho da pasta onde se encontra o ficheiro `satscan.exe` e o caminho onde o utilizador deseja que sejam guardados os ficheiros de *output* resultantes da análise. Neste último caminho é criada automaticamente uma pasta com a indicação do período temporal, sendo no seu interior criada, também de forma automática, uma pasta cujo nome referencia as AC dos casos analisados. É nesta segunda pasta que se encontra o ficheiro com os resultados da análise, o *output*.

A introdução de todos estes parâmetros é sequencial, ou seja, cada pedido de introdução surge após o anterior estar cumprido, seguindo-se uma figura ilustrativa. A resposta do utilizador encontra-se aqui a cor-de-laranja e a indicação por parte da ferramenta a cor-de-rosa.

```

Introduza a região que pretende analisar:continente|
Introduza o período temporal que pretende analisar:2013-2013|
Indique, face às AC escolhidas, o tipo de análise que pretende efetuar:g|
Indique a restante informação, face às AC, para a análise que pretende efetuar:50-56|
Introduza o caminho da pasta em que se encontram os ficheiros necessários para a análise:
C:\\Users\\afons\\Desktop\\Dados_Testar|
Introduza o caminho da pasta em que pretende guardar os resultados:C:\\Users\\afons\\Desktop|
Introduza o caminho da pasta em que se encontra o ficheiro SaTScan.exe:D:\\SaTScan|
Indique o valor e o critério de limitação do tamanho dos clusters, temporalmente:25-p|
Indique a percentagem máxima de nascimentos coberta por um cluster:25|

```

Figura 2.6: Colagem da introdução sequencial dos vários parâmetros durante o uso da ferramenta com exemplos de resposta.

2.3.2 Procedimentos incluídos na ferramenta

A ferramenta encontra-se em formato de ficheiro **R**, e efetua uma sequência de ações automatizadas que permitem a pesquisa facilitada de *clusters* espaço-temporais de AC.

Em primeiro lugar, os parâmetros são transformados em variáveis *string*. Em seguida, são lidos os

ficheiros de dados com a informação espaço-temporal dos casos, com a informação de quais as AC que os casos apresentam, com a informação espaço-temporal dos nascimentos em Portugal e com as coordenadas geográficas de interesse. Ficam guardadas quatro matrizes de dados, duas delas correspondentes à primeira e quarta folha do ficheiro do RENAC (Anexo 2), uma terceira com os dados dos nascimentos fornecidos pelo INE para os anos pretendidos, e uma quarta com as coordenadas geográficas dos concelhos para a região de Portugal escolhida. De notar que, embora os dados dos nascimentos sejam lidos em ficheiros separados por ano, estes são concatenados automaticamente numa única tabela, não sendo necessário, por isso, qualquer alteração ao ficheiro original, salvo a alteração do formato e nome do mesmo.

As quatro tabelas, já lidas, são filtradas de acordo com as opções do utilizador por forma a terem somente informações relevantes para a análise. Esta filtragem inclui a seleção dos casos que possuem as AC de interesse e dos nascimentos ocorridos apenas durante o período temporal indicado e na região de Portugal escolhida. Outras alterações são efetuadas, nomeadamente, para evitar incongruências quanto ao nome dos concelhos que impossibilitariam a análise, para terem as variáveis necessárias e na ordem correta que permita o uso destas tabelas no SaTScanTM, e incluírem a informação temporal com a data no formato ‘MM/AAAA’.

Mais concretamente, com a tabela que possui os códigos que identificam os casos na base de dados do RENAC e as suas respetivas AC, correspondente à quarta folha do ficheiro do RENAC, são guardados os códigos referentes aos indivíduos que possuem as AC escolhidas. Com esta informação, a ferramenta irá restringir a tabela que contém a informação espacial e temporal dos casos apenas aos portadores das AC de interesse.

Para obter os controlos, os casos são subtraídos ao total de nascimentos, por ano e mês de nascimento, e por concelho de residência da mãe. Os resultados destes passos são as três tabelas necessárias à utilização do SaTScanTM, uma com a informação espaço-temporal dos casos, outra com a informação espaço-temporal dos controlos, e a terceira com a informação das coordenadas geográficas dos concelhos. Estas três matrizes de dados são escritas em ficheiros em formato próprio para leitura no SaTScanTM.

Para terminar, são indicados os vários parâmetros da análise no `satscan.exe`: o nome dos ficheiros onde se encontram os dados, a data de início e de término da análise, que corresponderá ao primeiro dia do primeiro ano considerado e ao último dia do último ano indicado, respetivamente, o tipo de análise, e o modelo estatístico usado para a mesma, o nível de precisão dos dados e quaisquer restrições espaciais e temporais da janela de varrimento.

2.3.3 Interpretação do *output*

No ficheiro de resultados, cujo nome tem o formato ‘ano de início da análise’-‘ano de término da análise’.‘AC seleccionadas’.resultados.txt, há diversas informações relevantes:

- Cabeçalho: indica a data e hora em que foi efetuada a análise, qual o tipo de análise ('Retrospective Space-Time analysis'), quais os *clusters* de interesse ('scanning for clusters with high rates') e qual o modelo utilizado (neste caso, 'using the Bernoulli model').
- Resumo dos dados:
 1. 'Study Period': qual o período temporal analisado (ex: '2013/1/1 to 2013/12/31').
 2. 'Number of locations': número de localizações onde se encontram os dados analisados, neste caso, o número total de concelhos da região especificada (para Portugal continental são 278 municípios).
 3. 'Total population': indica o número total de nascimentos durante o período temporal analisado.
 4. 'Total number of cases': o número total de casos presentes na análise.
 5. 'Percent cases in area': a percentagem de casos na sua globalidade, ou seja, para todo o período temporal considerado e para toda a área em estudo.
- *Clusters*:
 1. 'Location IDs included': quais as localizações, neste caso os centroides dos municípios incluídos no *cluster*.
 2. 'Coordinates/radio': as coordenadas do centro do *cluster* e o raio, em quilómetros.
 3. 'Time frame': o período temporal de duração do *cluster* indicado pelas datas de início e de fim do mesmo.
 4. 'Population': número de nascimentos ocorridos nos municípios indicados durante o período temporal referido.
 5. 'Number of cases': número de casos de AC ocorridos nos municípios incluídos no *cluster* durante aquele período temporal.
 6. 'Expected cases': multiplicando o valor indicado em 'Percent cases in area' com o número de nascimentos no espaço-tempo deste *cluster* e dividindo posteriormente por 100, o valor apresentado será quantos casos haveria no *cluster* se a percentagem de casos no seu espaço-tempo fosse o mesmo da globalidade do período temporal e do conjunto de todas as localizações presentes na análise, ou seja, se H_0 for verdadeira.
 7. 'Observed/expected': o quociente entre o valor presente em 'Number of cases' e o valor em 'Expected cases'.
 8. 'Relative risk': o quociente entre o risco estimado no interior do *cluster* e o risco estimado fora do *cluster*; é calculado como o número de casos observados no interior do *cluster* dividido pelo número de casos esperados também no seu interior, dividido por sua vez pelo número de casos observados no exterior do *cluster* dividido pelo número de casos esperados igualmente no seu exterior.
 9. 'Percent cases in area': qual a percentagem de casos no período temporal e nas localizações do *cluster*.

10. ‘Log likelihood ratio’: valor obtido da estatística de teste.
11. ‘P-value’: *p-value* ajustado para testes múltiplos que advém da panóplia de janelas cilíndricas correspondentes a diferentes períodos temporais e localizações que constituem potenciais *clusters*, e que representa qual o *rank* do valor da estatística de teste do *cluster* com os valores obtidos da simulação de Monte Carlo.

Para ilustrar a descrição supra segue-se um exemplo, utilizando para isso o total de casos de AC do aparelho urinário (Q50-Q56) notificados ao RENAC durante o ano de 2013, em Portugal continental.

Program run on: Fri Oct 20 18:56:28 2017

Retrospective Space-Time analysis
scanning for clusters with high rates
using the Bernoulli model.

SUMMARY OF DATA

Study period.....: 2013/1/1 to 2013/12/31
Number of locations.....: 278
Total population.....: 78843
Total number of cases.....: 56
Percent cases in area.....: 0.07

CLUSTERS DETECTED

1.Location IDs included.: alpiarça
Coordinates / radius..: (39.241131 N, 8.570561 W) / 0 km
Time frame.....: 2013/4/1 to 2013/4/30
Population.....: 7
Number of cases.....: 2
Expected cases.....: 0.0050
Observed / expected...: 402.26
Relative risk.....: 417.12
Percent cases in area.: 28.6
Log likelihood ratio..: 10.351385
P-value.....: 0.032

2.Location IDs included.: nelas
Coordinates / radius..: (40.526873 N, 7.865207 W) / 0 km
Time frame.....: 2013/7/1 to 2013/7/31
Population.....: 7
Number of cases.....: 2
Expected cases.....: 0.0050
Observed / expected...: 402.26
Relative risk.....: 417.12
Percent cases in area.: 28.6
Log likelihood ratio..: 10.351385
P-value.....: 0.032

Figura 2.7: Porção inicial do ficheiro com os resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013 em Portugal continental.

Capítulo 3

Resultados e conclusões

Tendo em vista a demonstração da utilidade e potencialidade da ferramenta criada, foram escolhidos para as análises os casos com as anomalias indicadas na tabela seguinte, além de uma análise de todos os casos na sua globalidade. Esta seleção teve por base as AC que o EUROCAT considera mais importantes na vigilância epidemiológica (Loane, Dolk, Garne & Greenless, 2011). De notar que na análise dos grandes grupos de AC, indicados na tabela 2.1, são incluídos, nos grupos respetivos, os casos das AC mencionadas na tabela (2.2).

Tabela 3.1: As AC analisadas e o seu código segundo a CID-10.

AC analisadas	Código(s)
AC do Sistema Nervoso Central	Q00-Q07
Defeitos do tubo neural	Q00, Q01, Q05
Anencefalia	Q00
Encefalocelo	Q01
Microcefalia	Q02
Espinha bífida	Q05
AC do Olho, Ouvido, Face e Pescoço	Q10-Q18
AC do Aparelho Circulatório	Q20-Q28
Tronco arterial comum	Q20.0
Transposição dos grandes vasos	Q20.3
Tetralogia de Fallot	Q21.3
Coração esquerdo hipoplásico	Q23.4
Coartação da aorta	Q25.1
AC do Aparelho Respiratório	Q30-Q34
Fenda Labial e Fenda Palatina	Q35-Q37
Outras AC do Aparelho Digestivo	Q38-Q45
AC do Aparelho Genital	Q50-Q56
AC do Aparelho Urinário	Q60-Q64

AC e Deformações do Sistema Músculo-Esquelético	Q65-Q79
Polidactilia	Q69
Sindactilia	Q70
Hérnia diafragmática	Q79.0
Onfalocelo	Q79.2
Gastrosquisis	Q79.3
Outras AC	Q80-Q89
AC Cromossómicas não classificadas noutra local	Q90-Q99
Síndrome de Down	Q90
Síndrome de Edwards	Q91.0-Q91.3
Síndrome de Patau	Q91.4-Q91.7

3.1 Resultados de *clusters* espaço-temporais de AC para o ano de 2013

Nas tabelas seguintes encontram-se os resultados da análise de *clusters* espaço-temporais, para o ano de 2013, em Portugal continental, para as AC e grupos de AC indicadas na tabela 3.1. A escolha das análises se terem cingido ao ano de 2013 deve-se ao carácter ilustrativo das mesmas. Apresentam-se os *clusters* identificados como estatisticamente significativos, para um nível de significância $\alpha = 0,05$. Nas tabelas encontram-se as informações das coordenadas do centroide do *cluster*, o raio (em km) do mesmo, os concelhos nele incluídos, o intervalo temporal que ocupa, os nascimentos, casos ocorridos e casos esperados no *cluster*, e o *p-value* que lhe está associado. Todas as análises realizadas tiveram como parâmetros a limitação da componente temporal da janela máxima dos *clusters* de um mês e um máximo de 25% dos nascimentos da componente espacial da janela.

Não foram encontrados *clusters* espaço-temporais estatisticamente significativos, i.e. não foram encontradas diferenças no que diz respeito às prevalências, em espaço ou em tempo, de casos de: anencefalia, encefalocelo, AC do sistema circulatório, tronco arterial comum, transposição dos grandes vasos, tetralogia de Fallot, coração esquerdo hipoplásico, fendas labiais e palatinas, AC do aparelho urinário, polidactilia, hérnia diafragmática, onfalocelo, gastrosquisis, AC cromossómicas, síndrome de Down, síndrome de Edwards, e síndrome de Patau.

Tabela 3.2: Resultados da análise de *clusters* espaço-temporais de todos os casos de AC reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster nº1</i>	
Coordenadas do centro do <i>cluster</i>	40,519520 N, 7,245050 O
Raio do <i>cluster</i> (km)	38,55
Concelho(s)	Guarda, Celorico da Beira, Belmonte, Sabugal, Fornos de Algodres, Pínhel, Manteigas, Gouveia, Almeida, Trancoso, Aguiar da Beira, Penalva do Castelo
Intervalo temporal	2013/1/1 a 2013/1/31
Nascimentos	55
Casos	8
Casos esperados	0,75
<i>P-value</i>	0,010
<i>Cluster nº2</i>	
Coordenadas do centro do <i>cluster</i>	39,557917 N, 8,817010 O
Raio do <i>cluster</i> (km)	41,85
Concelho(s)	Porto de Mós, Batalha, Alcanena, Alcobaça, Nazaré, Torres Novas, Ourém, Santarém, Leiria, Rio Maior, Marinha Grande, Caldas da Rainha, Entroncamento, Golegã, Vila Nova da Barquinha, Tomar, Óbidos, Alpiarça, Cadaval
Intervalo temporal	2013/2/1 a 2013/2/28
Nascimentos	323
Casos	17
Casos esperados	4,41
<i>P-value</i>	0,036

Tabela 3.3: Resultados da análise de *clusters* espaço-temporais dos casos de AC do SNC (Q00-Q07) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	40,622305 N, 7,384944 O
Raio do <i>cluster</i> (km)	41,90
Concelho(s)	Celorico da Beira, Fornos de Algodres, Guarda, Trancoso, Gouveia, Aguiar da Beira, Penalva do Castelo, Pinhel, Manteigas, Mangualde, Sátão, Belmonte, Sernancelhe, Mêda, Seia, Almeida, Sabugal, Penedono, Nelas
Intervalo temporal	2013/1/1 a 2013/1/31
Nascimentos	93
Casos	5
Casos esperados	0,13
<i>P-value</i>	0,0026

Tabela 3.4: Resultados da análise de *clusters* espaço-temporais dos casos de defeitos do tubo neural (Q00,Q01,Q05) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	40,622305 N, 7,384944 O
Raio do <i>cluster</i> (km)	41,90
Concelho(s)	Celorico da Beira, Fornos de Algodres, Guarda, Trancoso, Gouveia, Aguiar da Beira, Penalva do Castelo, Pinhel, Manteigas, Mangualde, Sátão, Belmonte, Sernancelhe, Mêda, Seia, Almeida, Sabugal, Penedono, Nelas
Intervalo temporal	2013/1/1 a 2013/1/31
Nascimentos	93
Casos	3
Casos esperados	0,037
<i>P-value</i>	0,019

Tabela 3.5: Resultados da análise de *clusters* espaço-temporais dos casos de microcefalia (Q02) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,104890 N, 9,260428 O
Raio do <i>cluster</i> (km)	0
Concelho(s)	Torres Vedras
Intervalo temporal	2013/12/1 a 2013/12/31
Nascimentos	50
Casos	2
Casos esperados	0,0044
<i>P-value</i>	0,017

Tabela 3.6: Resultados da análise de *clusters* espaço-temporais dos casos de espinha bífida (Q05) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	40,622305 N, 7,384944 O
Raio do <i>cluster</i> (km)	41,90
Concelho(s)	Celorico da Beira, Fornos de Algodres, Guarda, Trancoso, Gouveia, Aguiar da Beira, Penalva do Castelo, Pinhel, Manteigas, Mangualde, Sátão, Belmonte, Sernancelhe, Mêda, Seia, Almeida, Sabugal, Penedono, Nelas
Intervalo temporal	2013/1/1 a 2013/1/31
Nascimentos	93
Casos	3
Casos esperados	0,022
<i>P-value</i>	0,0064

Tabela 3.7: Resultados da análise de *clusters* espaço-temporais dos casos de AC do olho, ouvido, face e pescoço (Q10-Q18) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	38,318208 N, 7,274933 O
Raio do <i>cluster</i> (km)	42,01
Concelho(s)	Mourão, Reguengos de Monsaraz, Moura, Barrancos, Alandroal, Portel, Vidigueira
Intervalo temporal	2013/9/1 a 2013/9/30
Nascimentos	34
Casos	3
Casos esperados	0,034
<i>P-value</i>	0,026

Tabela 3.8: Resultados da análise de *clusters* espaço-temporais dos casos de coartação da aorta (Q25.1) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	38,133426 N, 7,287442 O
Raio do <i>cluster</i> (km)	0
Concelho(s)	Moura
Intervalo temporal	2013/9/1 a 2013/9/30
Nascimentos	11
Casos	2
Casos esperados	0,0022
<i>P-value</i>	0,0073

Tabela 3.9: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho respiratório (Q30-Q34) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	40,688035 N, 7,905588 O
Raio do <i>cluster</i> (km)	0
Concelho(s)	Viseu
Intervalo temporal	2013/5/1 a 2013/5/31
Nascimentos	78
Casos	3
Casos esperados	0,037
<i>P-value</i>	0,025

Tabela 3.10: Resultados da análise de *clusters* espaço-temporais dos casos de AC do sistema digestivo não incluídas nos restantes grandes grupos (Q38-Q45) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	38,516130 N, 8,887487 O
Raio do <i>cluster</i> (km)	0
Concelho(s)	Setúbal
Intervalo temporal	2013/5/1 a 2013/5/31
Nascimentos	91
Casos	4
Casos esperados	0,12
<i>P-value</i>	0,045

Tabela 3.11: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster nº1</i>	
Coordenadas do centro do <i>cluster</i>	39,241131 N, 8,570561 O
Raio do <i>cluster</i> (km)	0
Concelho(s)	Alpiarça
Intervalo temporal	2013/4/1 a 2013/4/30
Nascimentos	7
Casos	2
Casos esperados	0,0050
<i>P-value</i>	0,032
<i>Cluster nº2</i>	
Coordenadas do centro do <i>cluster</i>	40,526873 N, 7,865207 O
Raio do <i>cluster</i> (km)	0
Concelho(s)	Nelas
Intervalo temporal	2013/7/1 a 2013/7/31
Nascimentos	7
Casos	2
Casos esperados	0,0050
<i>P-value</i>	0,032

Tabela 3.12: Resultados da análise de *clusters* espaço-temporais dos casos de AC do sistema músculo-esquelético (Q65-Q79) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,785241 N, 8,952636 O
Raio do <i>cluster</i> (km)	65,73
Concelho(s)	Marinha Grande, Leiria, Nazaré, Batalha, Alcobaça, Porto de Mós, Pombal, Ourém, Alcanena, Soure, Caldas da Rainha, Figueira da Foz, Torres Novas, Ansião, Alvaiázere, Rio Maior, Óbidos, Condeixa-a-Nova, Tomar, Santarém, Montemor-o-Velho, Entroncamento, Ferreira do Zêzere, Penela, Vila Nova da Barquinha, Bombarral, Golegã, Peniche, Figueiró dos Vinhos, Cadaval, Miranda do Corvo, Coimbra, Lourinhã
Intervalo temporal	2013/2/1 a 2013/2/28
Nascimentos	540
Casos	10
Casos esperados	1,50
<i>P-value</i>	0,035

Tabela 3.13: Resultados da análise de *clusters* espaço-temporais dos casos de sindactilia (Q70) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,187936 N, 8,082406 O
Raio do <i>cluster</i> (km)	58,06
Concelho(s)	Ponte de Sor, Avis, Abrantes, Chamusca, Mora, Alter do Chão, Constância, Gavião, Sousel, Crato, Fronteira, Coruche, Sardoal, Alpiarça, Vila Nova da Barquinha, Golegã, Almeirim, Entroncamento, Arraiolos, Mação, Nisa, Tomar, Torres Novas, Salvaterra de Magos, Vila de Rei, Estremoz, Monforte, Santarém
Intervalo temporal	2013/6/1 a 2013/6/30
Nascimentos	177
Casos	3
Casos esperados	0,049
<i>P-value</i>	0,031

Tabela 3.14: Resultados da análise de *clusters* espaço-temporais dos casos com outras AC não incluídas em qualquer outro grande grupo (Q80-Q89) reportados ao RENAC ocorridos em 2013 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,683118 N, 8,144312 O
Raio do <i>cluster</i> (km)	25,42
Concelho(s)	Vila de Rei, Sardoal, Ferreira do Zêzere, Sertã, Mação, Tomar, Proença-a-Nova
Intervalo temporal	2013/1/1 a 2013/1/31
Nascimentos	40
Casos	4
Casos esperados	0,034
<i>P-value</i>	0,00056

3.2 Resultados de *clusters* espaço-temporais de AC do aparelho circulatório (Q20-Q28)

Em seguida, apresentam-se os resultados da análise de *clusters* espaço-temporais para os casos de AC do sistema circulatório (Q20-Q28), para cada ano, desde 2000 a 2013. Os parâmetros utilizados são iguais aos da análise que originou os resultados supra.

O principal objetivo da exposição que se segue, é demonstrar especificamente a potencialidade do seguimento ao longo do tempo de *clusters* espaço-temporais, para poder estudar a sua evolução, ao invés da análise anterior, com a qual se pretendia ilustrar a ferramenta na sua globalidade para um determinado período de tempo.

A escolha do grupo das AC do sistema circulatório para esta análise deveu-se à sua elevada prevalência face aos restantes grupos para o período em estudo, de 2000 a 2013 (Braz et al., 2015). Não foram encontrados *clusters* espaço-temporais com significância estatística de casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC, ocorridos em Portugal continental, nos anos 2006 e 2013.

Tabela 3.15: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2000 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,854839 N, 7,501662 O
Raio do <i>cluster</i> (km)	123,11
Concelho(s)	Castelo Branco, Vila Velha de Ródão, Fundão, Proença-a-Nova, Oleiros, Idanha-a-Nova, Nisa, Castelo de Vide, Pampilhosa da Serra, Covilhã, Penamacor, Mação, Sertã, Marvão, Belmonte, Góis, Gavião, Vila de Rei, Arganil, Pedrógão Grande, Manteigas, Seia, Crato, Castanheira de Pêra, Sardoal, Oliveira do Hospital, Portalegre, Figueiró dos Vinhos, Tábua, Lousã, Sabugal, Ferreira do Zêzere, Gouveia, Abrantes, Alter do Chão, Vila Nova de Poiares, Miranda do Corvo, Penela, Alvaiázere, Guarda, Carregal do Sal, Santa Comba Dão, Ansião, Tomar, Nelas, Penacova, Constância, Mangualde, Arronches, Celorico da Beira, Fronteira, Vila Nova da Barquinha, Fornos de Algodres, Condeixa-a-Nova, Ponte de Sor, Mortágua, Coimbra, Monforte, Penalva do Castelo, Tondela, Ourém, Avis, Entroncamento, Mealhada, Torres Novas, Soure, Viseu, Almeida, Chamusca, Campo Maior, Pombal, Golegã, Sousel, Sátão, Aguiar da Beira, Anadia, Pinhel, Trancoso, Vouzela, Montemor-o-Velho, Elvas, Alcanena, Batalha, Águeda, Leiria, Cantanhede, Estremoz, Oliveira de Frades, Alpiarça, Vila Nova de Paiva, Borba, Oliveira do Bairro, Mora, Figueira da Foz, Porto de Mós, Sernancelhe, São Pedro do Sul, Santarém, Sever do Vouga, Almeirim, Vila Viçosa, Mêda, Castro Daire
Intervalo temporal	2000/2/1 a 2000/2/29
Nascimentos	1485
Casos	18
Casos esperados	4,98
<i>P-value</i>	0,035

Tabela 3.16: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2001 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	40,671306 N, 7,655881 O
Raio do <i>cluster</i> (km)	33,05
Concelho(s)	Penalva do Castelo, Mangualde, Sátão, Fornos de Algodres, Aguiar da Beira, Gouveia, Viseu, Celorico da Beira, Nelas, Vila Nova de Paiva, Serancelhe, Trancoso, Moimenta da Beira, Seia
Intervalo temporal	2001/6/1 a 2001/6/30
Nascimentos	194
Casos	7
Casos esperados	0,68
<i>P-value</i>	0,034

Tabela 3.17: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2002 em Portugal continental.

<i>Cluster n°1</i>	
Coordenadas do centro do <i>cluster</i>	39,692378 N, 8,574230 O
Raio do <i>cluster</i> (km)	38,87
Concelho(s)	Ourém, Batalha, Tomar, Alvaiázere, Torres Novas, Ferreira do Zêzere, Leiria, Porto de Mós, Alcanena, Entroncamento, Pombal, Vila Nova da Barquinha, Ansião, Marinha Grande, Golegã, Figueiró dos Vinhos, Vila de Rei, Constância, Alcobaça
Intervalo temporal	2002/9/1 a 2002/9/30
Nascimentos	470
Casos	10
Casos esperados	1,44
<i>P-value</i>	0,028
<i>Cluster n°2</i>	
Coordenadas do centro do <i>cluster</i>	37,151885 N, 8,728009 O
Raio do <i>cluster</i> (km)	170,95
Concelho(s)	Lagos, Portimão, Vila do Bispo, Aljezur, Monchique, Lagoa, Silves, Albufeira, Odemira, Loulé, Ourique, Almodôvar, Faro, São Brás de Alportel, Olhão, Sines, Castro Verde, Tavira, Santiago do Cacém, Aljustrel, Alcoutim, Mértola, Vila Real de Santo António, Castro Marim, Ferreira do Alentejo, Grândola, Beja, Alvito, Cuba, Alcácer do Sal, Serpa, Vidigueira, Viana do Alentejo, Setúbal, Sesimbra, Portel, Palmela, Seixal, Barreiro, Moura, Vendas Novas, Moita, Almada, Montemor-o-Novo, Évora
Intervalo temporal	2002/7/1 a 2002/7/31
Nascimentos	1329
Casos	16
Casos esperados	4,08
<i>P-value</i>	0,048

Tabela 3.18: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2003 em Portugal continental.

<i>Cluster nº1</i>	
Coordenadas do centro do <i>cluster</i>	37,125741 N, 8,453277 O
Raio do <i>cluster</i> (km)	48,06
Concelho(s)	Lagoa, Portimão, Silves, Albufeira, Monchique, Lagos, Aljezur, Loulé, Vila do Bispo, Faro
Intervalo temporal	2003/9/1 a 2003/9/30
Nascimentos	310
Casos	10
Casos esperados	0,77
<i>P-value</i>	0,000067
<i>Cluster nº2</i>	
Coordenadas do centro do <i>cluster</i>	41,460846 N, 8,315562 O
Raio do <i>cluster</i> (km)	16,69
Concelho(s)	Guimarães, Vizela, Braga, Fafe, Póvoa de Lanhoso, Felgueiras, Vila Nova de Famalicão
Intervalo temporal	2003/12/1 a 2003/12/31
Nascimentos	534
Casos	9
Casos esperados	1,32
<i>P-value</i>	0,048

Tabela 3.19: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2004 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,241131 N, 8,570561 O
Raio do <i>cluster</i> (km)	21,64
Concelho(s)	Alpiarça, Almeirim, Golegã, Chamusca, Santarém, Cartaxo
Intervalo temporal	2004/9/1 a 2004/9/30
Nascimentos	126
Casos	10
Casos esperados	0,35
<i>P-value</i>	$4,8 \times 10^{-7}$

Tabela 3.20: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2005 em Portugal continental.

<i>Cluster nº1</i>	
Coordenadas do centro do <i>cluster</i>	37,195684 N, 7,878642 O
Raio do <i>cluster</i> (km)	15,25
Concelho(s)	São Brás de Alportel, Tavira, Loulé
Intervalo temporal	2005/11/1 a 2005/11/30
Nascimentos	84
Casos	5
Casos esperados	0,20
<i>P-value</i>	0,0091
<i>Cluster nº2</i>	
Coordenadas do centro do <i>cluster</i>	39,407574 N, 9,088200 O
Raio do <i>cluster</i> (km)	18,15
Concelho(s)	Caldas da Rainha, Óbidos, Bombarral, Alcobaça
Intervalo temporal	2005/4/1 a 2005/4/30
Nascimentos	97
Casos	5
Casos esperados	0,23
<i>P-value</i>	0,019
<i>Cluster nº3</i>	
Coordenadas do centro do <i>cluster</i>	38,628427 N, 9,044965 O
Raio do <i>cluster</i> (km)	4,70
Concelho(s)	Barreiro, Moita
Intervalo temporal	2005/6/1 a 2005/6/30
Nascimentos	111
Casos	5
Casos esperados	0,26
<i>P-value</i>	0,037

Tabela 3.21: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2007 em Portugal continental.

<i>Cluster n°1</i>	
Coordenadas do centro do <i>cluster</i>	37,932941 N, 7,489637 O
Raio do <i>cluster</i> (km)	167,85
Concelho(s)	Serpa, Moura, Beja, Vidigueira, Mértola, Portel, Barrancos, Mourão, Cuba, Reguengos de Monsaraz, Castro Verde, Alcoutim, Alvito, Aljustrel, Ferreira do Alentejo, Castro Marim, Almodôvar, Évora, Viana do Alentejo, Alandroal, Ourique, Redondo, Tavira, Vila Real de Santo António, São Brás de Alportel, Loulé, Vila Viçosa, Santiago do Cacém, Borba, Alcácer do Sal, Grândola, Olhão, Arraiolos, Estremoz, Odemira, Faro, Silves, Montemor-o-Novo, Albufeira, Elvas, Sines, Sousel, Monchique, Vendas Novas, Mora, Lagoa, Monforte, Portimão, Campo Maior, Fronteira, Avis, Arronches, Aljezur, Montijo, Palmela, Setúbal, Lagos, Coruche, Alter do Chão, Ponte de Sor, Portalegre, Alcochete, Crato, Moita, Vila do Bispo, Benavente, Sesimbra, Barreiro, Seixal, Salvaterra de Magos, Marvão, Almeirim, Chamusca, Almada
Intervalo temporal	2007/4/1 a 2007/4/30
Nascimentos	1463
Casos	16
Casos esperados	3,44
<i>P-value</i>	0,0055
<i>Cluster n°2</i>	
Coordenadas do centro do <i>cluster</i>	39,552417 N, 8,990433 O
Raio do <i>cluster</i> (km)	38,04
Concelho(s)	Alcobaça, Nazaré, Porto de Mós, Caldas da Rainha, Batalha, Rio Maior, Marinha Grande, Óbidos, Alcanena, Leiria, Santarém, Bombarral, Cadaval, Peniche, Torres Novas
Intervalo temporal	2007/7/1 a 2007/7/31
Nascimentos	422
Casos	9
Casos esperados	0,99
<i>P-value</i>	0,0078

Tabela 3.22: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2008 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	39,271645 N, 8,380352 O
Raio do <i>cluster</i> (km)	52,36
Concelho(s)	Chamusca, Golegã, Alpiarça, Constância, Almeirim, Entroncamento, Vila Nova da Barquinha, Abrantes, Ponte de Sor, Torres Novas, Santarém, Alcanena, Salvaterra de Magos, Tomar, Coruche, Cartaxo, Sardoal, Rio Maior, Gavião, Avis, Mora, Azambuja, Porto de Mós, Ourém, Vila de Rei, Ferreira do Zêzere, Mação, Batalha
Intervalo temporal	2008/7/1 a 2008/7/31
Nascimentos	403
Casos	16
Casos esperados	1,07
<i>P-value</i>	$3,0 \times 10^{-9}$

Tabela 3.23: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2009 em Portugal continental.

<i>Cluster n°1</i>	
Coordenadas do centro do <i>cluster</i>	39,271645 N, 8,380352 O
Raio do <i>cluster</i> (km)	23,40
Concelho(s)	Chamusca, Golegã, Alpiarça, Constância, Almeirim, Entroncamento, Vila Nova da Barquinha
Intervalo temporal	2009/8/1 a 2009/8/31
Nascimentos	54
Casos	5
Casos esperados	0,13
<i>P-value</i>	0,0020
<i>Cluster n°2</i>	
Coordenadas do centro do <i>cluster</i>	39,819249 N, 8,395411 O
Raio do <i>cluster</i> (km)	34,90
Concelho(s)	Alvaiázere, Ferreira do Zêzere, Ansião, Figueiró dos Vinhos, Ourém, Penela, Pedrógão Grande, Tomar, Sertã, Vila de Rei, Pombal, Castanheira de Pêra, Miranda do Corvo, Condeixa-a-Nova, Soure, Leiria
Intervalo temporal	2009/9/1 a 2009/9/30
Nascimentos	278
Casos	7
Casos esperados	0,69
<i>P-value</i>	0,048

Tabela 3.24: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2010 em Portugal continental.

<i>Cluster n°1</i>	
Coordenadas do centro do <i>cluster</i>	39,854839 N, 7,501662 O
Raio do <i>cluster</i> (km)	140,53
Concelho(s)	Castelo Branco, V. Velha de Ródão, Fundão, Proença-a-Nova, Oleiros, Idanha-a-Nova, Nisa, Castelo de Vide, Pampilhosa da Serra, Covilhã, Penamacor, Mação, Sertã, Marvão, Belmonte, Góis, Gavião, V. de Rei, Arganil, Pedrógão Grande, Manteigas, Seia, Crato, Castanheira de Pêra, Sardoal, Oliveira do Hospital, Portalegre, Figueiró dos Vinhos, Tábua, Lousã, Sabugal, Ferreira do Zêzere, Gouveia, Abrantes, Alter do Chão, V. N. de Poiares, Miranda do Corvo, Penela, Alvaiázere, Guarda, Carregal do Sal, S. Comba Dão, Ansião, Tomar, Nelas, Penacova, Constância, Mangualde, Arronches, Celorico da Beira, Fronteira, V. N. da Barquinha, Fornos de Algodres, Condeixa-a-Nova, Ponte de Sor, Mortágua, Coimbra, Monforte, Penalva do Castelo, Tondela, Ourém, Avis, Entroncamento, Mealhada, Torres Novas, Soure, Viseu, Almeida, Chamusca, Campo Maior, Pombal, Golegã, Sousel, Sátão, Aguiar da Beira, Anadia, Pinhel, Trancoso, Vouzela, Montemor-o-Velho, Elvas, Alcanena, Batalha, Águeda, Leiria, Cantanhede, Estremoz, Oliveira de Frades, Alpiarça, V. N. de Paiva, Borba, Oliveira do Bairro, Mora, Figueira da Foz, Porto de Mós, Sernancelhe, S. Pedro do Sul, Santarém, Sever do Vouga, Almeirim, V. Viçosa, Méda, Castro Daire, Figueira de Castelo Rodrigo, Moimenta da Beira, Arraiolos, Marinha Grande, Mira, Vagos, Albergaria-a-Velha, Penedono, Aveiro, Vale de Cambra, Coruche, Tarouca, Ílhavo, Alcobaça, Rio Maior, Nazaré, Arouca, Cartaxo, Salvaterra de Magos, Estarreja, V. N. de Foz Côa, Redondo, Oliveira de Azeméis, Tabuaço, Alandroal, Armamar, Lamego, S. João da Pesqueira
Intervalo temporal	2010/1/1 a 2010/1/31
Nascimentos	1568
Casos	29
Casos esperados	4,65
<i>P-value</i>	$1,1 \times 10^{-9}$

<i>Cluster nº2</i>	
Coordenadas do centro do <i>cluster</i>	37,055635 N, 7,918243 O
Raio do <i>cluster</i> (km)	60,81
Concelho(s)	Faro, Olhão, São Brás de Alportel, Loulé, Tavira, Albufeira, Vila Real de Santo António, Castro Marim, Silves, Alcoutim, Lagoa, Almodôvar, Portimão
Intervalo temporal	2010/4/1 a 2010/4/30
Nascimentos	326
Casos	8
Casos esperados	0,97
<i>P-value</i>	0,045

Tabela 3.25: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2011 em Portugal continental.

<i>Cluster</i>	
Coordenadas do centro do <i>cluster</i>	40,121464 N, 7,482441 O
Raio do <i>cluster</i> (km)	106,79
Concelho(s)	Fundão, Covilhã, Belmonte, Castelo Branco, Penamacor, Manteigas, Seia, Pampilhosa da Serra, Oleiros, Idanha-a-Nova, Oliveira do Hospital, Gouveia, Arganil, Sabugal, Guarda, Tábua, Vila Velha de Ródão, Góis, Proença-a-Nova, Nelas, Celorico da Beira, Carregal do Sal, Mangualde, Fornos de Algodres, Castanheira de Pêra, Santa Comba Dão, Sertã, Penalva do Castelo, Pedrógão Grande, Lousã, Vila Nova de Poiares, Nisa, Penacova, Mação, Figueiró dos Vinhos, Tondela, Miranda do Corvo, Viseu, Castelo de Vide, Mortágua, Aguiar da Beira, Almeida, Sátão, Vila de Rei, Trancoso, Pinhel, Penela, Marvão, Coimbra, Vouzela, Sardoal, Gavião, Ferreira do Zêzere, Ansião, Alvaiázere, Mealhada, Condeixa-a-Nova, Vila Nova de Paiva, Sernancelhe, Anadia, Crato, Oliveira de Frades, Mêda, São Pedro do Sul, Águeda, Moimenta da Beira, Portalegre, Figueira de Castelo Rodrigo, Castro Daire, Tomar, Abrantes, Soure, Penedono, Sever do Vouga, Montemor-o-Velho, Oliveira do Bairro, Cantanhede, Tarouca, Alter do Chão, Constância, Pombal, Ourém, Vila Nova da Barquinha, Vila Nova de Foz Côa
Intervalo temporal	2011/3/1 a 2011/3/31
Nascimentos	921
Casos	29
Casos esperados	3,15
<i>P-value</i>	0,000015

Tabela 3.26: Resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho circulatório (Q20-Q28) reportados ao RENAC ocorridos em 2012 em Portugal continental.

<i>Cluster nº1</i>	
Coordenadas do centro do <i>cluster</i>	39,605177 N, 7,974881 O
Raio do <i>cluster</i> (km)	87,22
Concelho(s)	Mação, Sardoal, Vila de Rei, Proença-a-Nova, Gavião, Abrantes, Sertã, Nisa, Vila Velha de Ródão, Ferreira do Zêzere, Constância, Tomar, Oleiros, Vila Nova da Barquinha, Pedrógão Grande, Crato, Alvaiázere, Castelo de Vide, Figueiró dos Vinhos, Entroncamento, Ponte de Sor, Alter do Chão, Castelo Branco, Castanheira de Pêra, Torres Novas, Chamusca, Ourém, Golegã, Pampilhosa da Serra, Ansião, Góis, Penela, Marvão, Avis, Portalegre, Lousã, Alcanena, Miranda do Corvo, Alpiarça, Fronteira, Batalha, Arganil, Pombal, Condeixa-a-Nova, Fundão, Santarém, Almeirim, Vila Nova de Poiares, Porto de Mós, Leiria, Sousel, Soure, Mora, Monforte, Coimbra, Covilhã, Idanha-a-Nova, Penacova, Tábua, Arronches, Coruche, Rio Maior, Oliveira do Hospital, Marinha Grande, Cartaxo, Alcobaça, Salvaterra de Magos
Intervalo temporal	2012/10/1 a 2012/10/31
Nascimentos	864
Casos	20
Casos esperados	3,20
<i>P-value</i>	0,0000054
<i>Cluster nº2</i>	
Coordenadas do centro do <i>cluster</i>	40,791477 N, 7,334702 O
Raio do <i>cluster</i> (km)	31,95
Concelho(s)	Trancoso, Aguiar da Beira, Mêda, Celorico da Beira, Sernancelhe, Pinhel, Fornos de Algodres, Penedono, Sátão, Penalva do Castelo, Guarda, Moimenta da Beira
Intervalo temporal	2012/1/1 a 2012/1/31
Nascimentos	65
Casos	7
Casos esperados	0,24
<i>P-value</i>	0,000091

3.3 Conclusões

As análises de *clusters* espácio-temporais de AC cujos resultados foram apresentados correram de forma célere e ilustram a utilidade da ferramenta, podendo ser feito um grande número de análises num curto intervalo de tempo. Tendo em vista os objetivos já expressos, estas análises, de caráter exploratório, tiveram como parâmetros a limitação da componente espacial da janela de varrimento de uma cobertura geográfica máxima de 25% dos nascimentos em Portugal continental, para cada mês, e a limitação da componente temporal de um mês, para se encontrarem aglomerados pouco extensos, quer no tempo quer no espaço. No entanto, alguns resultados mostram que estas limitações podem, por vezes, não ser suficientes, particularmente a limitação percentual do número de nascimentos, dada a heterogeneidade demográfica que se verifica em Portugal e a desertificação do interior do território continental (Rosário, 2004). Por outras palavras, podem existir *clusters* muito extensos no espaço mas que cumpram a limitação da cobertura de 25% ou menos dos nascimentos em Portugal continental para o período temporal escolhido.

A juntar às diferenças no número de nascimentos, também o caráter facultativo da notificação de casos de AC ao RENAC pode contribuir para um enviesamento dos resultados, detetando-se potencialmente *clusters* de notificação de casos, e não *clusters* de aumento significativo do número de casos com significado epidemiológico onde se deva atuar, quer na prevenção, quer no estudo mais aprofundado das situações que os originaram. A deteção deste tipo de *clusters* mostra como a análise destes dados deve ser efetuada com alguma prudência, carecendo talvez de uma utilização prolongada da ferramenta a fim de se obter uma maior sensibilidade na determinação dos parâmetros a que obedecem as análises com vista a obter resultados de maior fiabilidade.

Não obstante, apesar dos resultados das análises realizadas com esta ferramenta serem afetados pelos dois fatores já mencionados, parece possuir relevância para o estudo das AC em Portugal. Por exemplo, parece pertinente uma pesquisa pormenorizada dos casos que ocorreram nos concelhos e nos meses dos *clusters*, a fim de se poder, porventura, identificar um fator comum que possa estar associado ao seu aparecimento. A metodologia de pesquisa de *clusters* espácio-temporais do SaTScanTM aqui exposta parece a mais indicada tendo em consideração a natureza dos dados analisados.

Será importante desenvolver uma investigação epidemiológica para confirmar ou excluir a existência de *clusters* sinalizados pela ferramenta aqui descrita.

Referências bibliográficas

Banta, J.E., Addison, A. & Beeson, W.L. (2015). Spatial patterns of epilepsy-related emergency department visits in California. *Journal of Public Health Research 2015 vol. 4, issue 1*: 441-449. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4407042/pdf/jphr-2015-1-441.pdf>.

Boyd, P.A., Haeusler, M., Barisic, I., Loane, M., Garne, E. & Dolk, H. (2011). Paper 1: The EURO-CAT network-organization and processes. Birth Defects Research Part A: *Clinical and Molecular Teratology vol. 91, Suppl 1: S2-15*. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/bdra.20780/full>.

Braz, P., Machado, A., Roquette, R. & Dias, C. M. (2015). Registo Nacional de Anomalias Congénitas: 11 anos de vigilância em Portugal. Lisboa: INSA, IP.

Carvalho, M., et. al. (2007a). Teratogénico. In *Enciclopédia Larousse* (Vol. XVII, pág. 6657). Santarém, Portugal: Temas e Debates Lda. e Larousse/VUEF.

Carvalho, M., et. al. (2007b). Nosocomial. In *Enciclopédia Larousse* (Vol. XIII, pág. 5069). Santarém, Portugal: Temas e Debates Lda. e Larousse/VUEF.

CDC – Centers for Disease Control and Prevention. (2016). Birth Defects Data and Statistics. Disponível em: <https://www.cdc.gov/ncbddd/birthdefects/data.html>. Consultado em fevereiro de 2017.

CDC – Centers for Disease Control and Prevention. (2017). Birth Defects Surveillance Toolkit. Disponível em: <https://www.cdc.gov/ncbddd/birthdefects/surveillancemanual/chapters/chapter-1/chapter1-4.html>. Consultado em dezembro de 2017.

Departamento de Epidemiologia. Instituto Nacional de Saúde Doutor Ricardo Jorge. (2011). Manual de apoio à codificação: Classificação Internacional de Doenças Capítulo XVII (Q00-Q99). Não publicado.

Dolk, H., Loane, M., Teljeur, C., Densem, J., Greenlees, R., McCullough, N., Morris, J., Nelen, V., Bianchi, F. & Kelly, A. (2015). Detection and investigation of temporal clusters of congenital anomaly in Europe: seven years of experience of the EUROCAT surveillance system. *European*

Journal of Epidemiology vol. 30, Issue 11: 1153-64. Disponível em: <http://link.springer.com/article/10.1007%2Fs10654-015-0012-y>.

Dolk, H., Loane, M. & Garne, E. (2010). The prevalence of congenital anomalies in Europe. *Advances in Experimental Medicine and Biology* vol. 686: 349-64. Disponível em: https://link.springer.com/chapter/10.1007%2F978-90-481-9485-8_20.

Ester, M., Kriegel, HP., Sander, J. & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). Disponível em: <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>.

EUROCAT (2009). EUROCAT Statistical Monitoring Protocol - 2009. EUROCAT Central Registry, University of Ulster. Disponível em https://ec.europa.eu/health/sites/health/files/rare_diseases/docs/eurocat_stat_2009_en.pdf. Consultado dezembro de 2016.

EUROCAT (*website*):

- Prevalence tables: Cases and prevalence (per 10,000 births) for all full member registries from 2010 to 2014. Disponível em: <http://www.eurocat-network.eu/accessprevalencedata/prevalencetables>. Consultado em fevereiro de 2017.

- Central Registry, University of Ulster. Disponível em http://www.eurocat-network.eu/content/Full%20Guide%201%204%20version%2008_Sept2017.pdf. Consultado em dezembro de 2016.

- What is EUROCAT?. Disponível em: <http://www.eurocat-network.eu/aboutus/whatiseurocat/whatiseurocat>. Consultado em agosto de 2017.

Glaz, J. & Zhang, Z. (2004). Multiple window discrete scan statistics. *Journal of Applied Statistics* vol. 31, issue 8: 967-980. Disponível em: <http://www.tandfonline.com/doi/pdf/10.1080/0266476042000270536>.

Han, J. & Kamber, M. (2001). *Data Mining Concepts and Techniques*. (2^a ed.). San Francisco, CA: Morgan Kaufmann Publishers. pp. 383-466.

Henry, K.A., Niu, X. & Boscoe, F.P. (2009). Geographic disparities in colorectal cancer survival. *International Journal of Health Geographics* 2009 vol.8: 48-61. Disponível em: <https://ij-healthgeographics.biomedcentral.com/track/pdf/10.1186/1476-072X-8-48?site=ij-healthgeographics.biomedcentral.com>.

INSA (*website*) RENAC. Disponível em: <http://www2.insa.pt/sites/INSA/Portugues/AreasCientificas/Epidemiologia/UnInstrObser/Paginas/RENAC.aspx>. Consultado em agosto de 2017.

Kim, J. H. & Scialli, A. R. (2011). Thalidomide: the tragedy of birth defects and the effective treatment of disease. *Toxicological Sciences* vol. 122, issue 1: 1-6. Disponível em: <https://academic.oup.com/toxsci/article/122/1/1/1672454/Thalidomide-The-Tragedy-of-Birth-Defects-and-the>.

Kleinman, K. (2015). rsatscan: Tools, Classes, and Methods for Interfacing with SaTScan Stand-Alone Software. R package version 0.3.9200. Disponível em: <https://CRAN.R-project.org/package=rsatscan>.

Kuehl, K.S. & Loffredo, C.A. (2006). A cluster of hypoplastic left heart malformation in Baltimore, Maryland. *Pediatric Cardiology* vol.27, issue 1: 25-31. Disponível em: <https://link.springer.com/article/10.1007%2Fs00246-005-0859-x>.

Kulldorff, M. & Nagarwalla, N. (1995). Spatial disease clusters: Detection and Inference. *Statistics in Medicine* vol. 14, issue 8: 799-810. Disponível em: <https://www.satscan.org/papers/k-sm1995.pdf>.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods* vol. 26, issue 6: 1481-1496. Disponível em: <https://www.satscan.org/papers/k-cstm1997.pdf>.

Kulldorff, M. (1999). Spatial scan statistics: Models, calculations and applications. *Recent Advances on Scan Statistics and Applications*: 302-322. Disponível em: <https://www.satscan.org/papers/k-scanbook1999.pdf>.

Loane, M., Dolk, H., Garne, E., Greenless, R. (2011). Paper 3: EUROCAT Data Quality Indicators for Population-Based Registries of Congenital Anomalies. *Birth Defects Research. Part A: Clinical and Molecular Teratology*, 91: S23-S30.

Nagarwalla, N. (1996). A scan statistic with a variable window. *Statistics in Medicine* vol. 15: 845-850. Disponível em: [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0258\(19960415\)15:7/9%3C845::AID-SIM254%3E3.0.CO;2-X/epdf](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0258(19960415)15:7/9%3C845::AID-SIM254%3E3.0.CO;2-X/epdf).

Naus, J.I. (1965). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association* vol. 60: 532-538. Disponível em: https://www.jstor.org/stable/2282688?seq=1#page_scan_tab_contents.

Ozdenerol, E., Williams, B. L., Kang, S.Y., Magsumbol, M.S. (2005). Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *International Journal of Health Geographics*: 4-19. Disponível em <https://ij-healthgeographics.biomedcentral.com/track/pdf/10.1186/1476-072X-4-19?site=ij-healthgeographics.biomedcentral.com>

Park, J., Yoon, S., Na, M.H., Song, H. (2015). The Effects of Air Pollution on Mortality in South Korea. *Procedia Environmental Sciences* vol. 26: 62-65. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1878029615001929>.

Patterson, D. (2009). Molecular genetic analysis of Down syndrome. *Human Genetics* vol. 126, issue 1: 195-214. Disponível em: <https://link.springer.com/article/10.1007%2Fs00439-009-0696-8>.

R Core Team (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Áustria. Disponível em: <https://www.R-project.org/>.

Rosário, L. (2004). Indicadores de Desertificação para Portugal Continental. Direção-Geral dos Recursos Florestais. Disponível em: <http://www2.icnf.pt/portal/pn/biodiversidade/ei/unccd-PT/pancd/resource/doc/dismed/DGRF%20Pub%20IndiceDesert-2004.pdf>

SaTScan (*website*). Software for the spatial, temporal and space-time statistics. Disponível em: <https://www.satscan.org>. Consultado de fevereiro a outubro de 2017.

Teljeur, C., Kelly, A., Loano, M., Densem, J., Dolk, H. (2015). Using scan statistics for congenital anomalies surveillance: the EUROCAT methodology. *European Journal of Epidemiology* vol. 30, Issue 11: 1165-73. <http://link.springer.com/article/10.1007%2Fs10654-015-0044-3>.

United Nations Development Programme (2013). Human Development Report 2013: Technical notes. Disponível em: http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf.

Viel, J.F., Floret, N., Mauny, F. (2005). Spatial and space-time scan statistics to detect low clusters of sex ratio. *Environmental and Ecological Statistics* vol. 12, issue 3: 289-299. Disponível em: <https://link.springer.com/article/10.1007/s10651-005-1514-7>

WHO – World Health Organization (2016). Congenital Anomalies Fact Sheet. Disponível em: <http://www.who.int/mediacentre/factsheets/fs370/en/>. Consultado em fevereiro de 2017.

Anexos

Anexo 1: Instrumento de notificação do RENAC

Mãe

Mãe		
<p>1 - Nome: _____ (Iniciais dos 4 primeiros nomes do B.I.)</p> <p>2 - Idade: _____ a a a a m m d d</p> <p>3 - Naturalidade: _____</p> <p>4 - Residência durante a gravidez: Distrito: _____ Concelho: _____ Freguesia: _____</p> <p>5 - Ocupação: _____</p> <p>6 - Etnia: _____</p> <p>7 - Gestações anteriores:</p> <p>7.1 - Abortos espontâneos: _____</p> <p>7.2 - Interrupções de Gravidez: _____</p> <p>7.3 - Fetos-mortos: _____</p> <p>7.4 - Nados vivos: _____</p> <p>7.5 - Total gestações anteriores: _____</p> <p>8 - Doenças crónicas: <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconhecido Especifique 1: _____ 2: _____</p>	<p>9 - Anomalias Congénitas: <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconhecido Especifique 1: _____ 2: _____</p> <p>10 - Reprodução medicamente assistida: <input type="checkbox"/> Sim <input type="checkbox"/> Não</p> <p>11 - Ocorrências durante esta gravidez:</p> <p>11.1 - Hábitos tabágicos no 1º Trim.: <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconhecido Especifique _____</p> <p>11.2 - Hábitos alcoólicos no 1º Trim.: <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconhecido Especifique _____</p> <p>11.3 - Toxicodependência no 1º Trim.: <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconhecido Especifique _____</p> <p>11.4 - Doenças no 1º Trimestre: _____ _____ _____</p> <p>11.5 - Medicamentos no 1º Trimestre: _____ _____ _____</p>	<p>11.6 - Ácido Fólico nesta gravidez: <input type="checkbox"/> Sim, início antes da gravidez _____ a a a a m m d d <input type="checkbox"/> Sim, início no 1º trimestre <input type="checkbox"/> Não tomou <input type="checkbox"/> Desconhecido</p> <p>11.7 - Outros agentes potencialmente teratogénicos: Especifique _____ _____</p> <p>12 - Familiares da mãe com anomalias: <input type="checkbox"/> Não <input type="checkbox"/> Não sabe <input type="checkbox"/> A mesma anomalia <input type="checkbox"/> Outra <input type="checkbox"/> A mesma e outra Especifique _____ _____</p> <p>13 - Estado migratório: <input type="checkbox"/> Imigrou de fora da UE durante a gravidez <input type="checkbox"/> Imigrou de fora da UE em adulta (≥ 18) <input type="checkbox"/> Imigrou de fora da UE com < 18 anos <input type="checkbox"/> Outra situação <input type="checkbox"/> Desconhecido</p>

Pai

Pai		
<p>1 - Idade: _____</p> <p>2 - Naturalidade: _____</p> <p>3 - Ocupação: _____</p> <p>4 - Etnia: _____</p> <p>5 - Anomalias congénitas: <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconhecido Especifique 1: _____ 2: _____</p>	<p>6 - Familiares do pai com anomalias: <input type="checkbox"/> Não <input type="checkbox"/> Não sabe <input type="checkbox"/> A mesma anomalia <input type="checkbox"/> Outra <input type="checkbox"/> A mesma e outra Especifique _____ _____</p>	<p>7 - Estado migratório: <input type="checkbox"/> Imigrou de fora da UE durante a gravidez <input type="checkbox"/> Imigrou de fora da UE em adulto (≥ 18) <input type="checkbox"/> Imigrou de fora da UE com < 18 anos <input type="checkbox"/> Outra situação <input type="checkbox"/> Desconhecido</p>

Consanguinidade dos Pais

Sim, especifique _____ Não Desconhecido

Descrição das Anomalias Presentes no Feto / RN

1 - _____

2 - _____

3 - _____

4 - _____

5 - _____

6 - _____

7 - _____

8 - _____

2 - Etiologia:

- | | |
|---------------------------------------|--|
| <input type="checkbox"/> Cromossómica | <input type="checkbox"/> Dominante (de novo) |
| <input type="checkbox"/> Familiar | <input type="checkbox"/> Outra genómica |
| <input type="checkbox"/> Isolada | <input type="checkbox"/> Síndrome |
| <input type="checkbox"/> Múltipla | <input type="checkbox"/> Teratogénica |

Comentários: _____

Anexo 2: Variáveis presentes na base de dados do RENAC

Variável	Descrição
Folha 1: “Ficha”	
Id	Código numérico de identificação da notificação na base de dados
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou recém-nascido (RN)
LocalNascimento	Local do nascimento (hospital, domicílio, etc.)
Sexo	Sexo do feto ou RN
Nascimento	Número de nascimentos resultantes da gravidez
Gestacao	O resultado da gestação (nado-vivo, feto morto, interrupção da gravidez, ou aborto espontâneo)
Peso	Peso à nascença do feto ou RN
NumMalformados	Número de fetos ou RN com AC, no caso de uma gravidez múltipla
UltimaMenstruacao	Data do último período menstrual da mãe
IdadeGestacional	Idade gestacional (IG) do feto ou RN ao nascer
MetodoIG	Método utilizado na determinação da IG
Identificacao1Anomalia	Quando foi identificada a primeira AC (diagnóstico pré-natal, ao nascer, etc.)
IG1Anomalia	Idade gestacional, em semanas, do feto quando foi detetada a primeira AC, se esta foi detetada em diagnóstico pré-natal
Estado1Anomalia	Estado do feto ou RN quando foi identificada a primeira AC
PrimeiroExameAlterado	Primeiro exame pré-natal alterado
OutrosTestes	Especificação de outros testes alterados
EcografiaObstetrica	Se, e quando, foram realizadas ecografias obstétricas
OutrosExamesMorfologicos	Outros exame morfológicos realizados
ColheitaProdutosFetais	Se, e por que motivo, foram realizadas colheitas de produtos fetais
TecnicaColheitaProdutosFetais	Técnica utilizada na colheita de produtos fetais
Cariotipo	Se foi realizado o cariótipo e qual o resultado do mesmo (normal, patológico...)
ResultadoCariotipo	Qual a fórmula do cariótipo
OutrosResultados	Resultados de outros testes realizados (imagiológicos, genéticos...)
SobreviveuSemana1	Se o RN sobreviveu à primeira semana após o nascimento
DataMorte	Data da morte do RN

Autopsia	Se foi realizada e se o seu resultado é conhecido
NomeMae	Iniciais do nome da mãe
IdadeMae	Idade da mãe
DataNascimentoMae	Data de nascimento da mãe
NaturalidadeMae	Naturalidade da mãe (país)
RegiaoMae	Região geográfica de residência da mãe
DistritoMae	Distrito de residência da mãe
ConcelhoMae	Concelho de residência da mãe
_FreguesiaMae	Freguesia de residência da mãe
CodigoPostal	Código postal de residência da mãe
OcupacaoMae	Ocupação da mãe
EtniaMae	Etnia da mãe
NumAbortosEspontaneos	Número de abortos espontâneos em gestações anteriores
NumInterrupcoesGravidez	Número de interrupções da gravidez em gestações anteriores
NumFetosMortos	Número de fetos mortos em gestações anteriores
NumNadosVivos	Número de nados vivos em gestações anteriores
NumGestacoes	Número total de gestações anteriores
DoencasCronicasMae	Diagnóstico médico de doenças crónicas da mãe
AnomaliasCongenitasMae	Se a mãe é portadora de AC
ReproducaoAssistida	Se a gravidez resulta de uma técnica de reprodução assistida
ReproducaoAssistidaTecnica	Qual a técnica de reprodução assistida utilizada
HabitosTabagicos	Hábitos tabágicos da mãe durante o 1º trimestre da gravidez
HabitosTabagicosQuais	Número de cigarros fumados por dia durante o 1º trimestre da gravidez
HabitosAlcoolicos	Hábitos alcoólicos da mãe durante o 1º trimestre da gravidez
HabitosAlcoolicosQuais	Frequência do consumo e tipo de bebida no 1º trimestre
Toxicodependencia	Consumo de droga durante o 1º trimestre
ToxicodependenciaQual	Estupefacientes consumidos no 1º trimestre
AcidoFolico	Se, e em que momento da gravidez, tomou ácido fólico
DataAcidoFolico	Data de início da toma de ácido fólico por parte da mãe
OutrosAgentesTeratogenicos	Outros agentes teratogénicos identificados
AnomaliasFamiliaresMae	Presença de AC em familiares da mãe
AnomaliasFamiliaresMaeQuais	Quais as AC presentes em familiares da mãe, e qual o grau de parentesco com o feto ou RN
EstadoMigratorioMae	Se a mãe imigrou de dentro ou fora da UE, e com que idade
IdadePai	Idade do pai
NaturalidadePai	Naturalidade do pai (país)
OcupacaoPai	Ocupação do pai
EtniaPai	Etnia do pai
AnomaliasCongenitasPai	Se o pai é portador de AC

AnomaliasFamiliaresPai	Se o pai tem familiares portadores de AC
AnomaliasFamiliaresPaiQuais	Quais as AC presentes em familiares do pai, e qual o grau de parentesco com o feto ou RN
EstadoMigratorioPai	Se o pai imigrou de dentro ou fora da UE, e com que idade
Consanguinidade	Se existe consanguinidade entre os pais
ConsanguinidadeQual	Qual o grau de consanguinidade
Etiologia	Qual a etiologia da(s) AC presentes no feto ou RN (isolada ou múltipla, cromossómica, familiar, etc.)
McKusick	Código que classifica as doenças de acordo com a sua hereditariedade
Comentario	Quaisquer comentários por parte do médico responsável pela notificação
Folha 2: “DoencasGravidezMae”	
Id	Número identificador de cada entrada na categoria respetiva (quando há mais que uma entrada, nesta folha, para um mesmo indivíduo, haverá diferentes “Id” relacionados com um mesmo “Id” da folha 1)
FichaId	Código numérico de identificação da notificação na base de dados (correspondente ao “Id” da folha 1; no caso de a mãe ter tido mais que uma doença, este número estará repetido)
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou RN
LocalNascimento	Local do nascimento (hospital, domicílio. . .)
Sexo	Sexo do feto ou RN
Peso	Peso à nascença do feto ou RN
Codigo	Código de acordo com a CID-10 da doença que mãe teve durante o 1º trimestre da gravidez
Nome	Nome da doença que a mãe teve durante o 1º trimestre da gravidez

Folha 3: “DoencasCronicasMae”	
Id	Número identificador de cada entrada na categoria respetiva (quando há mais que uma entrada, nesta folha, para um mesmo indivíduo, haverá diferentes “Id” relacionados com um mesmo “Id” da folha 1)
FichaId	Código numérico de identificação da notificação na base de dados (correspondente ao “Id” da folha 1; no caso de a mãe ter mais que uma doença crónica, este número estará repetido)
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou RN
LocalNascimento	Local do nascimento (hospital, domicílio. . .)
Sexo	Sexo do feto ou RN
Peso	Peso à nascença do feto ou RN
Codigo	Código de acordo com a CID-10 da doença crónica da mãe
Nome	Nome da doença crónica da mãe
Folha 4: “AnomaliasFeto”	
Id	Número identificador de cada entrada na categoria respetiva (quando há mais que uma entrada, nesta folha, para um mesmo indivíduo, haverá diferentes “Id” relacionados com um mesmo “Id” da folha 1)
FichaId	Código numérico de identificação da notificação na base de dados (correspondente ao “Id” da folha 1; no caso de o feto/RN ter mais que uma AC, este número estará repetido)
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou RN
LocalNascimento	Local do nascimento (hospital, domicílio. . .)
Sexo	Sexo do feto ou RN
Peso	Peso à nascença do feto ou RN
Codigo	Código de acordo com a CID-10 da AC observada no feto ou RN
Nome	Nome da AC presente no feto ou RN
Caminho	Identificação dos grupos hierárquicos da CID-10 onde se inclui a AC

Folha 5: “AnomaliasPai”	
Caminho	Identificação dos grupos hierárquicos da CID-10 onde se inclui a AC de que o pai é portador
FichaId	Código numérico de identificação da notificação na base de dados (correspondente ao “Id” da folha 1; no caso de o pai ter mais que uma AC, este número estará repetido)
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou RN
LocalNascimento	Local do nascimento (hospital, domicílio. . .)
Sexo	Sexo do feto ou RN
Peso	Peso à nascença do feto ou RN
Codigo	Código de acordo com a CID-10 da AC de que pai é portador
Nome	Nome da AC de que o pai é portador
Folha 6: “AnomaliasMae”	
Id	Número identificador de cada entrada na categoria respetiva (quando há mais que uma entrada, nesta folha, para um mesmo indivíduo, haverá diferentes “Id” relacionados com um mesmo “Id” da folha 1)
FichaId	Código numérico de identificação da notificação na base de dados (correspondente ao “Id” da folha 1; no caso de o pai ter mais que uma AC, este número estará repetido)
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou RN
LocalNascimento	Local do nascimento (hospital, domicílio. . .)
Sexo	Sexo do feto ou RN
Peso	Peso à nascença do feto ou RN
Codigo	Código de acordo com a CID-10 da AC de que a mãe é portadora
Nome	Nome da AC de que a mãe é portadora
Caminho	Identificação dos grupos hierárquicos da CID-10 onde se inclui a AC da mãe

Folha 7: “MedicamentosMae”	
Id	Número identificador de cada entrada na categoria respetiva (quando há mais que uma entrada, nesta folha, para um mesmo indivíduo, haverá diferentes “Id” relacionados com um mesmo “Id” da folha 1)
FichaId	Código numérico de identificação da notificação na base de dados (correspondente ao “Id” da folha 1; no caso de o pai ter mais que uma AC, este número estará repetido)
Centro	Centro hospitalar de onde provém a notificação
LocalRegisto	Código anonimizado do registo que permite, localmente, relacionar com o caso clínico
DataNascimento	Data de nascimento do feto ou RN
LocalNascimento	Local do nascimento (hospital, domicílio. . .)
Sexo	Sexo do feto ou RN
Peso	Peso à nascença do feto ou RN
Nome	Qual(ais) o(s) medicamento(s) tomado(s) pela mãe durante o 1º trimestre da gravidez

Anexo 3: *Script* integral da ferramenta

```
ac.esp.temp<-function(){

#Leitura das variáveis introduzidas pelo utilizador
reg<-readline("Introduza a região que pretende analisar:
              \n")
anoss<-readline("Introduza o período temporal que pretende analisar:
                \n")
anom1<-readline("Indique, face às AC escolhidas, o tipo de análise
                que pretende efetuar:
                \n")
anom2<-readline("Indique a restante informação, face às AC, para a
                análise que pretende efetuar:
                \n")
pat<-readline("Introduza o caminho da pasta em que se encontram os
              ficheiros necessários para a análise:
              \n")
dire<-readline("Introduza o caminho da pasta em pretende guardar os
               resultados:
               \n")
sats<-readline("Introduza o caminho da pasta em que se encontra o
               ficheiro SaTScan.exe:
               \n")
sst<-readline("Indique o valor e o critério de limitação do tamanho
              dos clusters, temporalmente:
              \n")
sss<-readline("Indique a percentagem máxima de nascimentos coberta
              por um cluster:
              \n")

#Transformação das variáveis introduzidas pelo utilizador para serem
              utilizadas
reg<-as.character(unlist(reg))
anoss<-as.character(unlist(anoss))
anom1<-as.character(unlist(anom1))
anom2<-as.character(unlist(anom2))
pat<-as.character(unlist(pat))
dire<-as.character(unlist(dire))
sats<-as.character(unlist(sats))
sst<-as.character(unlist(sst))
sss<-as.character(unlist(sss))
```

```

pat<-gsub("\\\\\\", "/" , pat , fixed=TRUE)

anos<-strsplit(anoss,"-")
anos<-unlist(anos)
sd<-as.numeric(anos[1])
ed<-as.numeric(anos[2])
anos<-seq(as.numeric(anos[1]),as.numeric(anos[2]))

#Leitura e transformação inicial dos ficheiros com os dados
cas1<-as.data.frame(read.delim(paste(pat,"/cas1.txt",sep=""),header=T
, stringsAsFactors = F))
cas2<-as.data.frame(read.delim(paste(pat,"/cas2.txt",sep=""),header=T
, stringsAsFactors = F))

pop<-as.data.frame(read.csv(paste(pat,"/pop",as.character(sd) ".csv",
sep=""), header=F, sep=";",
stringsAsFactors=FALSE, skip=8))
pop<-pop[-seq(nrow(pop)-9,nrow(pop)),]
pop[,1]<-rep(as.character(pop[1,1]),3756)
pop[,2]<-c(rep("01",313),rep("02",313),rep("03",313),rep("04",313),
rep("05",313),rep("06",313),
rep("07",313),rep("08",313),rep("09",313),rep("10",313),
rep("11",313),rep("12",313))

geo<-read.csv(paste(pat,"/geo.",reg,".txt",sep=""),header=T, sep=
",", stringsAsFactors=FALSE, encoding = "UTF-8")

#Caso o período temporal de análise compreenda mais que um ano
if (length(anos)>1){
anos<-anos[-1]
a<-c(pop[,1])
b<-c(pop[,2])
for (i in seq(1,length(anos))){
pop<-rbind(pop,read.csv(paste(pat,"/pop",as.character(anos[i]),".
csv",sep=""),
header=F, sep=";", stringsAsFactors=
FALSE, skip=8))
pop<-pop[-seq(nrow(pop)-9,nrow(pop)),]
for (i in seq(1,length(anos))){
a<-c(a,rep(as.character(anos[i]),3756))
b<-c(b,rep("01",313),rep("02",313),rep("03",313),rep("04",313),

```

```

      rep("05",313),rep("06",313),
      rep("07",313),rep("08",313),rep("09",313),rep("10",313),rep
      ("11",313),rep("12",313))}
  if (length(pop[,1])==length(a)){pop[,1]<-a}
  if (length(pop[,2])==length(b)){pop[,2]<-b}
}

pop<-pop[,-c(4,6,8)]

#Filtragem das tabelas com os dados
cas1<-data.frame(cbind(cas1$Id,cas1$ConcelhoMae,as.numeric(rep(1,nrow
  (cas1))),cas1$DataNascimento))
colnames(cas1)<-c("ID","Concelho","Casos","Data")
cas1$Data<-substr(cas1$Data,4,10)
cas1<-cas1[which(substr(cas1$Data,4,7) %in% as.character(seq(sd,ed)))
  ,]
cas2<-data.frame(cbind(as.character(cas2$FichaId),as.character(
  cas2$Codigo)))
colnames(cas2)<-c("ID","AC")

if (any(cas1$Concelho=="")==T){cas1<-cas1[-which(cas1$Concelho=="")
  ,]}
cas1$Concelho<-as.character(cas1$Concelho)
cas1$Concelho[intersect(which(cas1$Concelho=="Ilha de São Miguel"),
  which(cas1$Concelho=="Lagoa"))]<-"Lagoa (R.A.
  A.)"
cas1$Concelho[which(cas1$Concelho=="Vila Nova de Ourém")]<-"Ourém"
cas1$Concelho[which(cas1$Concelho=="Alpiarça")]<-"Alpiarça"
cas1$Concelho[which(cas1$Concelho=="Meda")]<-"Mêda"
cas1$Concelho<-tolower(cas1$Concelho)
cas1$Concelho<-gsub(" ","_",cas1$Concelho)

geo<-data.frame(cbind(geo[,2],geo[,5],geo[,4]))
colnames(geo)<-c("Concelho","Latitude","Longitude")
geo$Concelho<-tolower(geo$Concelho)
geo$Concelho<-gsub(" ","_",geo$Concelho)
if (reg=="madeira"){geo$Concelho[which(geo$Concelho=="calheta")]<-"
  calheta_(r.a.m.)"}
if (reg=="açores.cen"){geo$Concelho[which(geo$Concelho=="
  praia_da_vitória")]<-"vila_da_praia_da_vitória"}
if (reg=="açores.cen"){geo$Concelho[which(geo$Concelho=="
  santa_cruz_da_graciosa")]<-

```

```

    "santa_cruz_da_graciosa_(r.a.a.)"}
if (reg=="açores.cen"){geo$Concelho[which(geo$Concelho=="velas")]<-"
    velas_(r.a.a.)"}
if (reg=="açores.cen"){geo$Concelho[which(geo$Concelho=="
    calheta_de_são_jorge")]<-"calheta_(r.a.a.)"}
if (reg=="açores.ori"){geo$Concelho[which(geo$Concelho=="lagoa")]<-"
    lagoa_(r.a.a.)"}

colnames(pop)[3]<-"Concelho"
pop<-pop[-which(pop[3]=="Portugal"),]
pop<-pop[-which(pop[3]=="Continente"),]
pop<-pop[-which(pop[3]=="Região Autónoma dos Açores"),]
pop<-pop[-which(pop[3]=="Região Autónoma da Madeira"),]
pop<-pop[-which(pop[3]=="Ignorado"),]
pop$Concelho<-tolower(pop$Concelho)
pop$Concelho<-gsub(" ", "_", pop$Concelho)

date<-c()
for (i in seq(1:nrow(pop))) {
    date<-c(date, paste(as.character(pop[i,2]), as.character(pop[i,1]),
        sep="-"))
}

pop<-data.frame(cbind(pop[3], pop[4], pop[5], date))
colnames(pop)<-c("Concelho", "FM", "NV", "Data")

pop$NV<-as.numeric(pop$NV)
pop$FM<-as.numeric(pop$FM)
pop$NT<-pop$NV+pop$FM

pop<-pop[,-c(2,3)]
pop<-pop[,c(1,3,2)]

cas1<-cas1[which(cas1$Concelho %in% geo$Concelho),]
pop<-pop[which(pop$Concelho %in% geo$Concelho),]

cas2$ID<-as.character(cas2$ID)
cas2$AC<-as.character(cas2$AC)

#Seleção dos casos consoante as escolhas das AC a analisar
if (anom1=="g") {
    anom2<-anom2

```

```

anom2<-strsplit (anom2," -")
anom2<-unlist (anom2)
anom<-seq (as.numeric (anom2 [1] ) ,as.numeric (anom2 [2] ) )
ac<-as.character (anom)
for (i in seq (1,length (ac))) {
  if (nchar (ac [i] )==1){
    ac [i]<-paste ("Q0" ,ac [i] ,sep="") }
  else {ac [i]<-paste ("Q" ,ac [i] ,sep="") }}
if (anom [1]==10 & anom [length (anom)]==18){ac<-c (ac ," D18.10" ," K07
  .0" ," K07.1" ) }
if (anom [1]==20 & anom [length (anom)]==28){ac<-c (ac ," P83.2" ) }
if (anom [1]==38 & anom [length (anom)]==45){ac<-c (ac ," P75" ) }
if (anom [1]==65 & anom [length (anom)]==79){ac<-c (ac ," K40" ," K42" ," P96
  .3" ) }
if (anom [1]==80 & anom [length (anom)]==89){ac<-c (ac ," C41.4" ," C49.0" ,"
  C49.4" ," D18.00" ," D18.01" ," D82.1" ) }}

if (anom1=="s" ) {
  anom22<-anom2
  anom2<-strsplit (anom2," -")
  ac<-unlist (anom2)
  for (i in seq (1,length (ac))) {
    ac [i]<-paste ("Q" ,ac [i] ,sep="") }}

if (anom1=="e" ) {
  ac<-anom2}

if (anom1=="g" ) {
  acq<-ac [ which (substr (ac ,1 ,1)=="Q" ) ]
  acn<-ac [ which (substr (ac ,1 ,1)!="Q" ) ]
  idq<-unique (cas2$ID [ which (substr (cas2$AC ,1 ,3) %in% acq) ])
  idn<-unique (cas2$ID [ which (cas2$AC %in% acn) ])
  id<-unique (c (idq ,idn) )
  cas1<-cas1 [ which (cas1$ID %in% id) ,]
  cas1<-cas1 [, -1]}

if (anom1=="s" ) {
  id1<-nchar (ac [1] )
  id<-unique (cas2$ID [ which (substr (cas2$AC ,1 ,id1) %in% ac) ])
  cas1<-cas1 [ which (cas1$ID %in% id) ,]
  cas1<-cas1 [, -1]}

```

```

if (anom1=="e") {
  id1<-nchar(ac)
  id<-unique(cas2$ID[which(substr(cas2$AC,1,id1) == ac)])
  cas1<-cas1[which(cas1$ID %in% id),]
  cas1<-cas1[, -1]}

if (anom1=="t") {cas1<-cas1[, -1]}

if (anom1=="t") {anomref<-"t"}
if (anom1=="g") {anomref<-anom22}
if (anom1=="s") {anomref<-anom22}
if (anom1=="e") {anomref<-ac}

install.packages("plyr")
library("plyr")

cas1$Casos<-as.numeric(cas1$Casos)
cas1<-ddply(cas1,.(Concelho,Data),summarize,Casos=sum(Casos))

#Tabela com os controlos
ctl<-as.data.frame(pop)
colnames(ctl)[2]<-"CTL"

ctlcas<-merge(ctl,cas1,by=c("Concelho","Data"),all=T)
ctlcas$Casos[is.na(ctlcas$Casos)==T]<-0
ctlcas$CTL<-ctlcas$CTL-ctlcas$Casos
ctl<-ctlcas[, -c(4)]

cas<-ctlcas[,c(1,4,2)]
ctl<-ctl[,c(1,3,2)]
ctl$CTL[which(ctl$CTL<0)]<-0

#Determinação de tudo o que é necessário para o rsatscan
install.packages("rsatscan")
library(rsatscan)
require(rsatscan)

sats<-gsub("\\\\\\\\", "/", sats, fixed=TRUE)
dire<-gsub("\\\\\\\\", "/", dire, fixed=TRUE)
dire1<-paste(dire, "/", anoss, sep="")
dir.create(dire1)
dire2<-paste(dire1, "/", anomref, sep="")

```

```

dir.create(dire2)
dire<-dire2

nomest<-paste(anoss , anomref , sep="")
nomest<-gsub("-", "", nomest , fixed=T)
nomecas<-paste(nomest , ". cas" , sep="")
nomectl<-paste(nomest , ". ctl" , sep="")
nomegeo<-paste(nomest , ". geo" , sep="")
nomeout<-paste(gsub("/", "\\\\" , dire))

sst<-unlist(strsplit(sst , "-"))
sst2<-sst[1]
if(sst[2]=="m"){sst1<-"1"}
if(sst[2]=="p"){sst1<-"0"}
sd<-as.character(sd)
ed<-as.character(ed)

invisible(ss.options(reset=TRUE))

#Parâmetros de input dos dados no SaTScan
ss.options(list(CaseFile=nomecas))
ss.options(list(ControlFile=nomectl))
ss.options(list(PrecisionCaseTimes=2))
ss.options(paste("StartDate=",sd,"/1/1",sep=""))
ss.options(paste("EndDate=",ed,"/12/31",sep=""))
#Population file não usado
ss.options(list(CoordinatesFile=nomegeo))
ss.options("UseGridFile=n")
#Grid file não usado
ss.options("CoordinatesType=1")
#Parâmetros da análise de clusters no SaTScan
ss.options(list(AnalysisType=3)) #Retrospective Space-Time
ss.options(list(ModelType=1)) #Bernoulli
ss.options("ScanAreas=1") #Pesquisa de áreas com maior prevalência
ss.options(list(TimeAggregationUnits=2)) #Mês
ss.options(list(TimeAggregationLength=1)) #1(mês)
#Parâmetros do ficheiro de output
ss.options("ResultsFile=D:\\DS\\paula.braz\\Desktop\\sdfgs.txt")
ss.options("OutputGoogleEarthKML=n")
ss.options("OutputShapefiles=n")
ss.options("MostLikelyClusterEachCentroidASCII=y")
ss.options("MostLikelyClusterEachCentroidDBase=n")

```

```

ss.options("MostLikelyClusterCaseInfoEachCentroidASCII=y")
ss.options("MostLikelyClusterCaseInfoEachCentroidDBase=n")
ss.options("CensusAreasReportedClustersASCII=y")
ss.options("CensusAreasReportedClustersDBase=n")
ss.options("IncludeRelativeRisksCensusAreasASCII=y")
ss.options("IncludeRelativeRisksCensusAreasDBase=n")
ss.options("SaveSimLLRsASCII=y")
ss.options("SaveSimLLRsDBase=n")
#Limitação do limite máximo da componente espacial da janela
ss.options(paste("MaxSpatialSizeInPopulationAtRisk=",sss,sep=""))
#Limitação do limite máximo da componente temporal da janela
ss.options(paste("MaxTemporalSizeInterpretation=",sst1,sep=""))
ss.options(paste("MaxTemporalSize=",sst2,sep=""))

#Escrita dos ficheiros necessários ao SaTScan nos formatos requeridos
write.ss.prm(dire,nomest)

write.cas(cas,dire,nomest)
write.ctl(ctl,dire,nomest)
write.geo(geo,dire,nomest)

#Análise propriamente dita e escrita do ficheiro de output
out<-satscan(dire,nomest,sslocation = sats)
out<-lapply(out,write,paste(dire2,"/",sd,"-",ed,".",anomref,".
    resultados.txt",sep=""),append=T)
}

```

Anexo 4: Instruções de utilização da ferramenta de análise de *clusters* espaço-temporais

A ferramenta criada realiza a análise de *clusters* espaço-temporais para cada uma das regiões de Portugal listadas abaixo, dentro de um intervalo temporal pretendido, para o grupo de AC definidas pelo utilizador (podem ser todas). Abaixo serão descritas as variáveis que devem ser fornecidas pelo utilizador e qual a forma de introdução das mesmas. De notar que nos exemplos fornecidos, estes devem ser sempre considerados como não tendo plicas. Estes parâmetros são indicados de forma sequencial na mesma ordem que são apresentados abaixo.

Juntamente com o ficheiro que contém ferramenta, serão fornecidos ao RENAC os ficheiros com as coordenadas geográficas dos vários municípios de Portugal e os ficheiros com os valores dos nascimentos por mês e por concelho, desde o ano 2000 a 2013, todos nos respetivos formatos prontos a serem lidos e usados nas análises.

1. Região.

Para seleccionar a região para qual vai ser realizada a análise deve introduzir-se a respetiva designação, que se encontra do lado direito da lista seguinte:

- Continente: ‘continente’;
- Madeira: ‘madeira’;
- Açores Ocidental: ‘açores.oci’;
- Açores Central: ‘açores.cen’;
- Açores Oriental: ‘açores.ori’;

2. Período temporal.

Para definir o período temporal da análise, devem ser escritos o ano de início e o ano de término separados por um hífen, utilizando-se a mesma regra para analisar um único ano, como se exemplifica de seguida: para analisar o ano 2000, será necessário introduzir ‘2000-2000’, e para analisar o período de 2000 a 2009, deve ser escrito ‘2000-2009’.

3. Modo de escolha das AC a analisar.

Este item compreende quatro possibilidades distintas:

- ‘t’, para analisar todos casos de AC;
- ‘g’, para analisar apenas os casos de AC de um dos grandes grupos apresentados na tabela 2.1;
- ‘s’, para analisar os casos de um conjunto de AC específico;
- ‘e’, para analisar os casos de uma determinada AC;

4. Escolha das AC a analisar.

Para cada hipótese no item acima, o modo de introdução de quais as AC a incluir na análise, é:

- ‘t’: deve ser reintroduzido ‘t’;
- ‘g’: devem ser inseridos os limites superior e inferior do grande grupo pretendido, separados por hífen, sem o carácter ‘Q’; a título ilustrativo, para analisar os casos de AC do SNC, deve ser introduzido ‘00-07’; nos casos em que se aplicar, os casos com AC compreendidas na tabela 2.2 serão incluídos nos grupos respetivos;
- ‘s’: devem ser escritos todos os códigos das AC que se pretende incluir na análise, sem o carácter ‘Q’: separados por hífen, tendo em atenção que todos os códigos indicados devem possuir o mesmo número de caracteres; exemplificando, para analisar os casos de AC com síndrome de Edwards, deve ser escrito ‘91.0-91.1-91.2-91.3’;
- ‘e’: deve ser escrito o código da AC cujos casos se pretende analisar; a título ilustrativo, para analisar os casos de AC com sindactilia será escrito ‘Q70’.

5. Caminho da pasta com as tabelas necessárias.

Este parâmetro serve para indicar o caminho onde se encontra a pasta que contém as tabelas necessárias ao funcionamento da ferramenta, nomeadamente:

- Ficheiro de texto com os valores separados por tabulações correspondente à folha “Ficha” da base de dados do RENAC, com a coluna ‘Comentarios’ eliminada; o seu nome deve ser: ‘cas1.txt’.
- Ficheiro de texto com os valores separados por tabulações correspondente à folha “Anomalias” da base de dados do RENAC; o seu nome deve ser: ‘cas2.txt’.
- Ficheiro com os valores separados por ponto e vírgula contendo a informação do INE sobre os nados-vivos e óbitos fetais por mês e por concelho, em Portugal, com os valores separados por semi-vírgulas (*comma-separated values*), para o(s) ano(s) relevante(s); o seu nome deve ser: ‘pop#####.csv’, representado os ‘#’ aqui o ano a que o ficheiro se refere; antes dos ficheiros serem guardados em formato de *comma-separated values* para serem usados pela ferramenta, deverão ter um aspeto similar ao da figura que se segue.

	A	B	C	D	E	F	G	H
1								
2	Quadro extraído em 18 de Maio de 2017 (15:45:23)							
3	http://www.ine.pt							
4								
5								
6	Período de referência dos dados (1)	Lista de meses	Local de residência da mãe (NUTS - 2002) (2)		Óbitos fetais (N.º) por Local de residência da mãe (NUTS - 2002) e Mês (óbito); Anual	Nados-vivos (N.º) por Local de residência da mãe (NUTS - 2002), Sexo e Mês de nascimento; Anual (3)		
7						Sexo		
8						HM		
9						N.º		
10						N.º		
10	2000	Janeiro	Portugal	PT	45	9825		
11			Continente	1	43	9292		
12			Arcos de V	1111601	0	12		
13			Caminha	1111602	0	12		
14			Melgaço	1111603	0	5		
15			Monção	1111604	0	12		
16			Paredes de	1111605	0	10		
17			Ponte da B	1111606	0	13		
18			Ponte de Li	1111607	0	37		
19			Valença	1111608	0	11		
20			Viana do C	1111609	0	75		

Figura A.1: Exemplo do ficheiro com os valores totais de nascimentos para cada concelho de Portugal, por mês, referentes ao ano 2000.

- Ficheiro de texto com os valores separados por vírgulas com as coordenadas geográficas dos centroides dos concelhos de Portugal para a região sobre a qual se deseja realizar a análise; o seu nome deverá ser: 'geo."REGIÃO".csv', substituindo-se "REGIÃO" pela designação da região a que diz respeito o ficheiro, com a codificação do primeiro parâmetro apresentado da ferramenta; e.g. o ficheiro com as coordenadas geográficas dos centroides dos concelhos do grupo ocidental do arquipélago dos Açores deve ter o nome 'geo.açores.oci.csv'; as variáveis incluídas devem ser, por esta ordem: numeração da linha, a começar no 0 a partir da primeiro município; nome do município; código DDCC do município; longitude; latitude.

O conteúdo da pasta que contém os ficheiros mencionados será então semelhante ao apresentado na figura seguinte.

Nome	Tipo	Tamanho
cas1	Documento de texto	4 KB
cas2	Documento de texto	4 KB
geo.açores.cen	Documento de texto	1 KB
geo.açores.oci	Documento de texto	1 KB
geo.açores.ori	Documento de texto	1 KB
geo.continente	Documento de texto	17 KB
geo.madeira	Documento de texto	1 KB
pop2000	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2001	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2002	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2003	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2004	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2005	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2006	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2007	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2008	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	110 KB
pop2009	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	109 KB
pop2010	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	109 KB
pop2011	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	109 KB
pop2012	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	109 KB
pop2013	Ficheiro de Valores Separados Por Vírgulas do Microsoft Excel	109 KB

Figura A.2: Imagem de uma pasta com todos os ficheiros prontos a ser lidos pela ferramenta.

O caminho deve ser escrito com duas contrabarras ('\\') a separar cada diretório, ao invés de apenas uma, que é o mais usual. A título ilustrativo: 'C:\\Users\\afons\\Desktop\\Dados'.

6. Caminho da pasta onde se pretende guardar os ficheiros com os resultados.

Aqui deve ser introduzido o caminho onde se pretende que sejam guardados os ficheiros resultantes da análise. Através do uso da ferramenta são criadas automaticamente duas pastas, estando a segunda contida na primeira, no caminho que for indicado:

- primeira pasta: o nome que esta toma corresponde ao período de estudo da análise efetuada; por exemplo, caso esteja a ser analisado o período do ano 2013, o nome da pasta criada será ‘2013-2013’;
- segunda pasta (inserida na primeira): é criada no interior da primeira e o seu nome corresponde às AC escolhidas para análise; exemplificando, se forem analisados os casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013, no interior da primeira pasta denominada ‘2013-2013’, encontrar-se-à a segunda pasta ‘50-56’; é no interior desta pasta que serão escritos de forma automática os ficheiros necessários à utilização do SaTScanTM, no formato adequado e com as alterações necessárias já efetuadas, bem como o ficheiro que contém os resultados da análise, o ficheiro de *output*.

Os ficheiros usados para realizar a análise com o SaTScanTM mencionados supra, e ilustrados na figura A.3, são:

- um ficheiro com os parâmetros (com a extensão ‘.prm’);
- um ficheiro com a tabela referente aos casos (com a extensão ‘.cas’);
- um ficheiro com a tabela referente aos controlos (com a extensão ‘.ctl’);
- um ficheiro com a tabela referente à informação geográfica (com a extensão ‘.geo’).

O ficheiro de *output* terá o nome, para o caso já descrito, ‘2013-2013.50-56.resultados.txt’.

O caminho do diretório onde se pretende que sejam criadas estas duas pastas deve ser escrito como o anterior, com duas contrabarras a separar cada diretório. A título ilustrativo: ‘C:\\Users\\afons\\Desktop\\Análises_20_Out_Continente’. As pastas criadas serão semelhantes às que se apresentam em seguida.

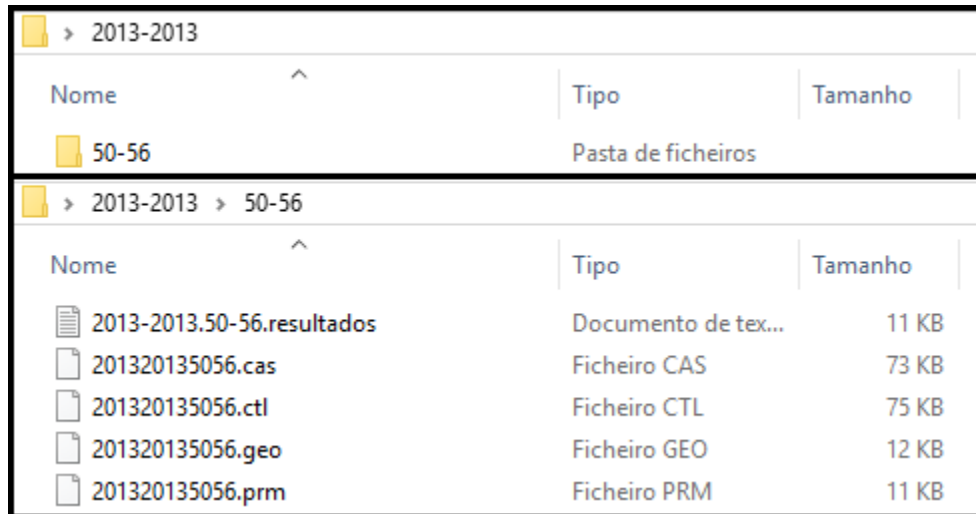


Figura A.3: (Em cima) imagem da primeira pasta criada, cujo nome indica os anos do período considerado para a análise e que contém a segunda pasta; (em baixo) imagem da segunda pasta criada, no interior da primeira, cujo nome indica as AC escolhidas para a análise.

7. Caminho da pasta que contém o ficheiro `satscan.exe`.

Para responder a este parâmetro, deve ser introduzido o caminho onde se encontra o ficheiro `satscan.exe`, escrevendo também aqui duas contrabarras entre diretórios. Por exemplo: 'F:\\SaTScan'.

8. Limite máximo temporal.

Este parâmetro irá afetar diretamente a análise efetuada, limitando o máximo que a componente temporal dos *clusters* encontrados apresentam. Pode limitar-se por número de meses ou por percentagem do período analisado. A título ilustrativo, se é pretendido detetar *clusters* com um mês de duração temporal máxima, deve ser introduzido '1-m'; se é pretendido detetar *clusters* com uma duração temporal máxima de 25% do período analisado, deve ser introduzido '25-p'.

9. Limite máximo espacial.

À semelhança do parâmetro anterior, este será o limite máximo, agora da componente espacial, que os *clusters* encontrados na análise apresentam. A limitação é realizada por percentagem máxima de nascimentos, casos e controlos, incluídos na componente espacial dos *clusters* detetados. Por exemplo, para serem reportados apenas *clusters* contendo no máximo 25% dos nascimentos ocorridos no período temporal dos mesmos, será necessário escrever '25'.

O ficheiro de *output*

Os ficheiros de *output* da análise possuem diversos resultados e informações, descritos de seguida com exemplos das análises efetuadas no decorrer deste trabalho:

- Cabeçalho: indica a data e hora em que foi efetuada a análise, qual o tipo de análise ('Retrospective Space-Time analysis'), quais os *clusters* de interesse ('scanning for clusters with high rates') e qual o modelo utilizado (neste caso, 'using the Bernoulli model').
- Resumo dos dados:
 1. 'Study Period': qual o período temporal analisado (ex: '2013/1/1 to 2013/12/31').
 2. 'Number of locations': número de localizações onde se encontram os dados analisados, neste caso, o número total de concelhos da região especificada (para Portugal continental são 278 municípios).
 3. 'Total population': indica o número total de nascimentos durante o período temporal analisado.
 4. 'Total number of cases': o número total de casos presentes na análise.
 5. 'Percent cases in area': a percentagem de casos na sua globalidade, ou seja, para todo o período temporal considerado e para toda a área em estudo.
- *Clusters*:
 1. 'Location IDs included': quais as localizações, neste caso os centroides dos municípios incluídos no *cluster*.
 2. 'Coordinates/radio': as coordenadas do centro do *cluster* e o raio, em quilómetros.
 3. 'Time frame': o período temporal de duração do *cluster* indicado pelas datas de início e de fim do mesmo.
 4. 'Population': número de nascimentos ocorridos nos municípios indicados durante o período temporal referido.
 5. 'Number of cases': número de casos de AC ocorridos nos municípios incluídos no *cluster* durante aquele período temporal.
 6. 'Expected cases': multiplicando o valor indicado em 'Percent cases in area' com o número de nascimentos no espaço-tempo deste *cluster* e dividindo posteriormente por 100, o valor apresentado será quantos casos haveria no *cluster* se a percentagem de casos no seu espaço-tempo fosse o mesmo da globalidade do período temporal e do conjunto de todas as localizações presentes na análise, ou seja, se H_0 for verdadeira.
 7. 'Observed/expected': o quociente entre o valor presente em 'Number of cases' e o valor em 'Expected cases'.

8. ‘Relative risk’: o quociente entre o risco estimado no interior do *cluster* e o risco estimado fora do *cluster*; é calculado como o número de casos observados no interior do *cluster* dividido pelo número de casos esperados também no seu interior, dividido por sua vez pelo número de casos observados no exterior do *cluster* dividido pelo número de casos esperados igualmente no seu exterior.
9. ‘Percent cases in area’: qual a percentagem de casos no período temporal e nas localizações do *cluster*.
10. ‘Log likelihood ratio’: valor obtido da estatística de teste.
11. ‘P-value’: *p-value* ajustado para testes múltiplos que advém da panóplia de janelas cilíndricas correspondentes a diferentes períodos temporais e localizações que constituem potenciais *clusters*, e que representa qual o *rank* do valor da estatística de teste do *cluster* com os valores obtidos da simulação de Monte Carlo.

Para ilustrar a descrição supra segue-se um exemplo, utilizando para isso o total de casos de AC do aparelho urinário (Q50-Q56) notificados ao RENAC durante o ano de 2013, em Portugal continental.

Program run on: Fri Oct 20 18:56:28 2017

Retrospective Space-Time analysis
scanning for clusters with high rates
using the Bernoulli model.

SUMMARY OF DATA

Study period.....: 2013/1/1 to 2013/12/31
Number of locations.....: 278
Total population.....: 78843
Total number of cases.....: 56
Percent cases in area.....: 0.07

CLUSTERS DETECTED

1.Location IDs included.: alpiarça
Coordinates / radius..: (39.241131 N, 8.570561 W) / 0 km
Time frame.....: 2013/4/1 to 2013/4/30
Population.....: 7
Number of cases.....: 2
Expected cases.....: 0.0050
Observed / expected...: 402.26
Relative risk.....: 417.12
Percent cases in area.: 28.6
Log likelihood ratio..: 10.351385
P-value.....: 0.032

2.Location IDs included.: nelas
Coordinates / radius..: (40.526873 N, 7.865207 W) / 0 km
Time frame.....: 2013/7/1 to 2013/7/31
Population.....: 7
Number of cases.....: 2
Expected cases.....: 0.0050
Observed / expected...: 402.26
Relative risk.....: 417.12
Percent cases in area.: 28.6
Log likelihood ratio..: 10.351385
P-value.....: 0.032

Figura A.4: Porção inicial do ficheiro com os resultados da análise de *clusters* espaço-temporais dos casos de AC do aparelho genital (Q50-Q56) reportados ao RENAC ocorridos em 2013 em Portugal continental.