

IDENTIFICAÇÃO DE CONGLOMERADOS ESPACIAIS DE ACIDENTES AÉREOS NO BRASIL¹

Nara Núbia Vieira²
André Luiz Fernandes Cançado³

Artigo submetido em 09/01/2013

Aceito para publicação em 15/03/2013

RESUMO: O presente artigo tem por objetivo verificar a existência de conglomerados espaciais de acidentes aéreos nacionais a fim de identificar regiões onde a proporção de acidentes, considerando o total de ocorrências, seja significativamente maior que no restante do Brasil. Com esse resultado, a orientação da prevenção de acidentes se torna mais precisa, aumentando consequentemente a Segurança de Voo no país. Para tal, utilizou-se a estatística de Varredura de Kulldorff, adequada para análise de agrupamentos espaciais. Foram consideradas as ocorrências aeronáuticas civis que aconteceram entre os anos de 2007 e 2011 dentro dos limites do território brasileiro em que o município ou a coordenada geográfica do local do evento foram identificados.

PALAVRAS-CHAVE: Análise de Agrupamento. Acidente Aéreo. Varredura Circular. Conglomerado Espacial. Segurança de Voo.

1 INTRODUÇÃO

Atualmente, o Brasil apresenta um crescimento significativo na aviação, o que pode ser comprovado pelo aumento da frota de aeronaves registrada na ANAC, pelo número de movimentos em aeroportos controlados pela INFRAERO e pelo volume de combustível de aviação vendido, divulgado pela ANP. Esse desenvolvimento aumenta a exposição ao risco gerando perdas consideráveis de vidas humanas e de patrimônio público e privado. Diante desse cenário na aviação brasileira, observa-se que há uma demanda por estudos que orientem a prevenção de acidentes aeronáuticos.

¹ Artigo derivado do trabalho de conclusão de curso da pós-graduação em Estatística Aplicada da Universidade Cruzeiro do Sul.

² Bacharel em Estatística pela Universidade de Brasília (2008) e Pós-Graduada em Estatística Aplicada pela Universidade Cruzeiro do Sul (2013). Atua na área de prevenção de acidentes aeronáuticos desde 2010.

³ Bacharel em Matemática Computacional (2003), Mestre em Estatística (2005) e Doutor em Engenharia Elétrica (2009), com ênfase em Ciência da Computação, todos pela Universidade Federal de Minas Gerais. É professor do Departamento de Estatística da Universidade de Brasília desde 2010. Tem experiência na área de Estatística Computacional e Ciência da Computação, atuando principalmente em problemas de Estatística Espacial e Otimização.

Para fazer este estudo estatístico sobre ocorrências aeronáuticas, foram utilizados dados do Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA). Esse Centro é o órgão central do Sistema de Investigação e Prevenção de Acidentes Aeronáuticos (SIPAER) o qual tem a responsabilidade legal de investigar ocorrências aeronáuticas com finalidade única de evitar recorrências, e não de encontrar e punir possíveis culpados (BRASIL, 1982).

A técnica estatística utilizada neste trabalho foi a Varredura Circular de Kulldorff (1997). Esse método identifica *clusters* espaciais por meio de um processo iterativo que varre o mapa à procura de conglomerados, partindo de uma ocorrência e incluindo, em cada passo, as ocorrências mais próximas. O objetivo é verificar se há uma região no país em que a probabilidade de ocorrer um acidente é significativamente maior do que fora dela. Para verificar essa significância foram utilizadas simulações de Monte Carlo.

Uma vez que o crescimento da aviação é fator essencial para o desenvolvimento do país, e levando em conta que cada vez mais pessoas procuram esses serviços, esse trabalho visa contribuir com a Segurança de Voo, já que será possível identificar locais que se tornam mais críticos por apresentarem uma maior probabilidade de ocorrência de acidentes, orientando assim as ações preventivas.

2 REVISÃO BIBLIOGRÁFICA

Segundo os dados do CENIPA (BRASIL, 2012), o número de acidentes aéreos no país sobe desde 2005, com um crescimento mais acentuado a partir de 2007. De 58 acidentes em 2005 passou-se a 102 em 2007 e no ano de 2011 foram 159 (ver Figura 1). Esse aumento é atribuído a diversos fatores como falta de fiscalização, aumento das atividades aéreas, entre outros. Diante disso, surge a necessidade de orientar as atividades de prevenção que têm como objetivo fortalecer a Segurança de Voo.

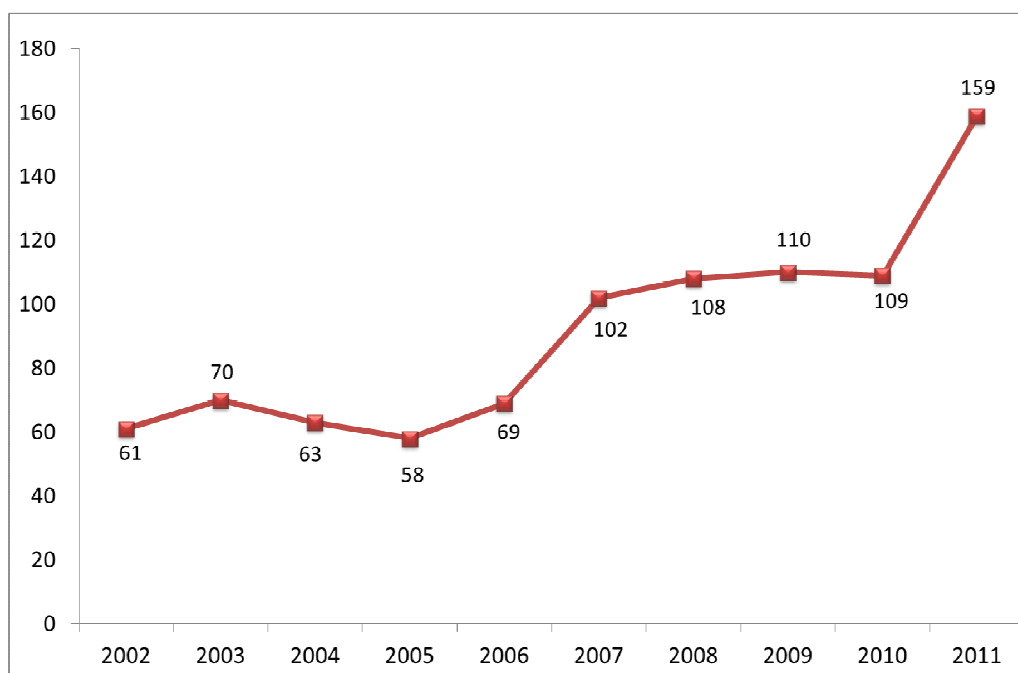


FIGURA 1 - Total de Acidentes da Aviação Civil Brasileira por Ano. Fonte: CENIPA (2012).

As ocorrências aeronáuticas podem ser diferenciadas entre acidentes, incidentes graves e incidentes. Essa conceituação está na Norma do Sistema do Comando da Aeronáutica (NSCA) 3-1, de 31 de outubro de 2008 (BRASIL, 2008).

Entre 2007 e 2011, no território brasileiro, aconteceram 2.714 ocorrências com aeronaves civis que não estão classificadas como “experimentais” no Registro Brasileiro de Aeronaves Civis (RBAC). Desse total, foram excluídas 43 ocorrências, pois o município ou coordenada geográfica do local do evento não foram identificados.

3 PREVENÇÃO DE ACIDENTES AERONÁUTICOS

Herbert William Heinrich foi um dos precursores da engenharia de segurança e ganhou notoriedade com seu estudo sobre acidentes industriais. Por meio de dados estatísticos, verificou que para cada acidente que causou uma lesão grave, havia 29 ocorrências com lesões leves e 300 situações de perigo (WIKIPEDIA, 2012).

Essa teoria foi de grande importância para que os profissionais da área de Segurança de Voo se dedicassem a estudos de ocorrências aeronáuticas de menor repercussão (incidentes e incidentes graves), pois, diminuindo esses

casos, o número de acidentes também tenderia a cair.

O mapa da Figura 2 mostra os locais onde aconteceram as ocorrências aeronáuticas civis entre os anos de 2007 e 2011. O tamanho dos pontos é proporcional ao número de ocorrências daquele ponto. Os círculos pretos representam apenas os Acidentes enquanto os azuis representam todas as Ocorrências (acidentes, incidentes graves e incidentes).

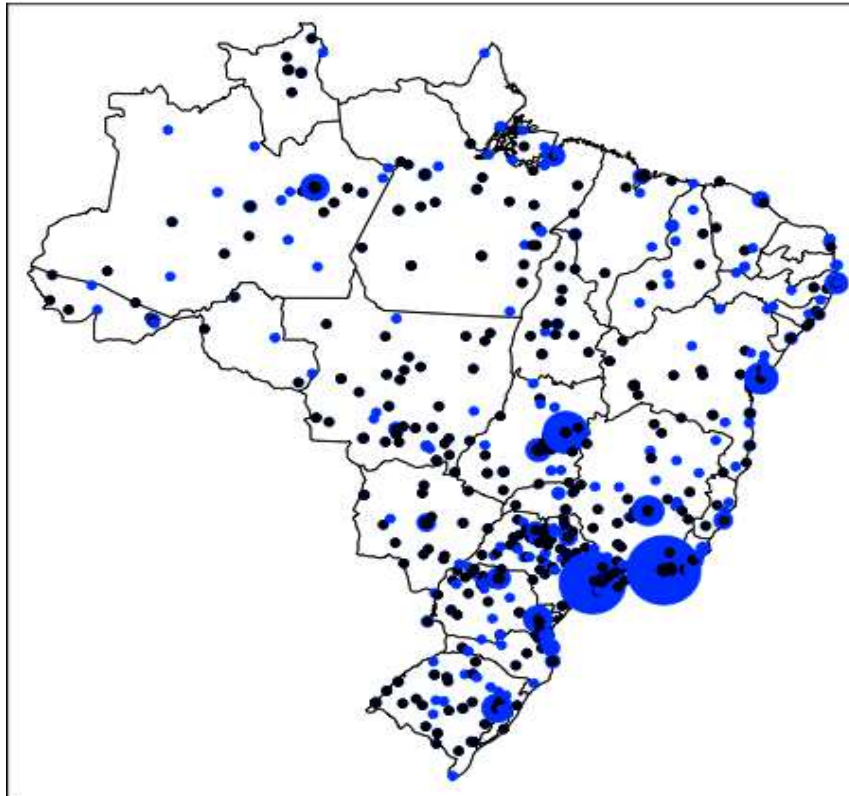


FIGURA 2 - Distribuição das Ocorrências de 2007 a 2011 da Aviação Civil Brasileira.
Fonte: CENIPA (2012).

4 METODOLOGIA

Os dados de acidentes aéreos são fornecidos pelo CENIPA e foram considerados para este estudo aqueles que ocorreram com aeronaves civis não experimentais, entre os anos de 2007 e 2011, dentro dos limites do território brasileiro. Do total de 2.714 acidentes encontrados, foram desconsiderados 43, pois as coordenadas geográficas do local do evento ou o município não foram identificados.

A estatística espacial de Varredura Circular foi proposta por Kulldorff (1997) para identificar conglomerados espaciais. Este método consiste em um processo iterativo que varre o mapa através de janelas circulares e, por meio

de um teste de razão de verossimilhança, verifica se os conglomerados são significativos ou não.

Para aplicar esta técnica, é necessário que se tenham, disponíveis, as coordenadas geográficas de cada ocorrência, aqui denotadas por *casos*. Além disso, é necessário, ainda, que se tenham as coordenadas geográficas dos eventos que não são consideradas casos, chamados *controles*. A união dos casos e dos controles constitui a *população*. No presente estudo, consideramos população a contagem total de acidentes, incidentes graves e incidentes, sendo que os casos de interesse são somente os acidentes. As coordenadas dos eventos foram as fornecidas pelo IBGE, de acordo com o município da ocorrência aeronáutica.

4.1 Teste de Hipóteses e Estatística do Teste

Assumindo que a contagem de eventos segue, em cada região, uma distribuição Poisson, queremos verificar se a proporção de acidentes em relação ao número total de ocorrências em uma determinada área é significativamente maior do que fora dela. Dessa forma, a função que descreve a probabilidade de ocorrência de c casos na i -ésima região, n_i , é:

$$f_i(c) = \begin{cases} \frac{e^{-\lambda_i} \lambda_i^c}{c!}, & c \geq 0 \\ 0, & c < 0 \end{cases} \quad (3.1)$$

onde λ_i é o número esperado de casos na i -ésima região, proporcional à sua população.

Chamaremos de *zona* um subconjunto de regiões. Para uma zona z , sejam n_z e c_z a população e o número de casos em z , respectivamente, e p a probabilidade de que um elemento qualquer da população (acidentes, incidentes graves e incidentes) nessa zona seja um caso (acidente). O número esperado de casos na zona z , μ_z , é proporcional à sua população, ou seja, $\mu_z = pn_z$. Logo, temos que a função de probabilidade de acidentes nessa zona é:

$$f_z(c) = \begin{cases} \frac{\mu_z^{c_z} e^{-\mu_z}}{c_z!}, & c \geq 0 \\ 0, & \text{c.c} \end{cases} \tag{3.2}$$

Considerando a população total como N , e C como o número total de casos, e ainda que n_z e c_z sejam a população e o número de casos na zona z , a população e o número de casos fora da zona z são dados, respectivamente, por

$$\begin{aligned} n_{\bar{z}} &= N - n_z \\ c_{\bar{z}} &= C - c_z \end{aligned}$$

Sob a hipótese nula de que não há *clusters* no território brasileiro, a probabilidade de que ocorra um acidente dentro de uma zona qualquer z deve ser igual à de ocorrer fora dela. Seja p a probabilidade de que ocorra um caso dentro desta zona e q a de que ocorra fora. Então estamos interessados no seguinte teste de hipóteses:

$$\begin{aligned} H_0) & p = q \\ H_1) & p > q \end{aligned}$$

Para testar as hipóteses acima, Kulldorff [6] usou como estatística de teste a razão de verossimilhança para o modelo Poisson, descrita abaixo. Os detalhes desse cálculo podem ser encontrados em seu artigo. A razão de verossimilhança é dada por:

$$LR(z) = \frac{L}{L_0} = \begin{cases} \left(\frac{c_z}{\mu_z}\right)^{c_z} \left(\frac{C - c_z}{C - \mu_z}\right)^{C - c_z}, & c_z > \mu_z \\ 1, & \text{c.c} \end{cases} \tag{3.3}$$

O cluster mais verossímil pode ser encontrado pela maximização do logaritmo da equação 3.3 sobre todas as possíveis zonas. Dessa forma, temos a seguinte estatística de teste:

$$T = \max_z \log(LR(z)) \tag{3.4}$$

4.2 Identificação dos Clusters

A busca pelo cluster mais verossímil é feita por um método iterativo que obtém uma sequência de zonas candidatas formadas por regiões próximas. Ao final do processo, a zona z^* correspondente ao maior valor de razão de verossimilhança será a zona mais verossímil. O processo se inicia calculando a distância euclidiana entre os pares de regiões segundo suas coordenadas geográficas (x_i, y_i) , fornecidas pelo IBGE para o município da ocorrência aeronáutica. Então, a distância entre dois municípios i e j é:

$$D_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

As distâncias entre todos os m municípios, nos quais houve pelo menos um caso estudado, formam uma matriz simétrica quadrada com m colunas e m linhas. Portanto, $d_{i,j}$ é o elemento da i -ésima linha e j -ésima coluna dessa matriz:

$$d_{i,j} = \begin{cases} D_{i,j}, & i \neq j \\ 0, & \text{c.c} \end{cases}$$

Assim, temos a matriz de distâncias:

$$D = \begin{bmatrix} 0 & D_{1,2} & \cdots & D_{1,j} & \cdots & D_{1,m} \\ D_{2,1} & 0 & \cdots & D_{2,j} & \cdots & D_{2,m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ D_{i,1} & D_{i,2} & \cdots & 0 & \cdots & D_{i,m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ D_{m,1} & D_{m,2} & \cdots & D_{m,j} & \cdots & 0 \end{bmatrix}$$

Com as distâncias calculadas, o processo de Varredura Circular inicia-se. Considere $D_{i,j}$ a distância entre as regiões r_i e r_j . Fixando-se a região j e ordenando o vetor das distâncias resultante (a j -ésima coluna da matriz D), em ordem crescente, temos que $D_{(0),j}$ é a distância entre a região r_j e a i -ésima região mais próxima de r_j .

$$\begin{bmatrix} 0 \\ D_{(2),1} \\ D_{(3),1} \\ \vdots \\ D_{(m),1} \end{bmatrix}$$

Partindo de $t = 1$, a primeira zona considerada será composta apenas pela região r_1 , ou seja, $z_1 = \{r_1\}$. Dessa forma, calcula-se pela equação 3.3 a estatística $T = LR(z_1)$. A segunda zona é composta pela região r_1 e por aquela que está à distância $D_{(2),1}$ por ser a mais próxima dela. Essa zona é representada por $z_2 = \{r_1, r_{(2)}\}$ e, segundo a mesma fórmula, é possível calcular o valor de $LR(z_2)$. Se este valor superar o atual valor da estatística T , esta terá seu valor alterado para $LR(z_2)$. Caso contrário, mantém-se o valor corrente de T .

O método continua agregando novas regiões por ordem de distância, sempre comparando o resultado da estatística T atual com o novo valor, mantendo-se o maior deles. Recomenda-se que os conglomerados candidatos tenham no máximo 50% da população. Logo, o processo de agregação cessa quando a população da zona candidata atinge esse limite.

Esse procedimento é repetido iniciando-se em todas as demais regiões (isto é, para $j = 2, 3, \dots, m$) até encontrar a zona mais verossímil z^* correspondente ao maior valor da equação 3.3.

De forma resumida, o procedimento varre o mapa com janelas circulares de centros e raios variados, comparando a razão entre o número observado e esperado de casos dentro e fora da janela. Ao final, a zona escolhida como solução mais verossímil será a que apresentar maior discrepância na incidência de casos, levando-se em conta o número de casos e a população da zona.

Além do conglomerado mais verossímil, chamado de conglomerado primário, o procedimento pode reportar ainda conglomerados secundários que não tenham interseção entre si. Assim, o conglomerado secundário será o conglomerado mais verossímil que não possui interseção com o primário. O conglomerado terciário será o conglomerado mais verossímil que não possui

interseção com o primário nem com o secundário, e assim por diante.

4.3 Inferência Estatística

A verificação da significância dos clusters é feita através de simulações de Monte Carlo, já que a distribuição da estatística T sob H_0 é desconhecida. Essa simulação consiste em distribuir no mapa, repetidas vezes, o número total de casos C aleatoriamente sob a hipótese nula. Dessa forma, a probabilidade de uma ocorrência ser um acidente é igual em qualquer parte do mapa e assume o valor $p = C/N$. Suponha que o número de simulações seja s . Para cada simulação, isto é, para cada distribuição aleatória dos casos sob a hipótese nula, será calculada a estatística T , conforme o procedimento descrito na seção 3.2, e o conjunto de todos os valores T_1, \dots, T_s formará uma distribuição empírica da estatística do teste (FIGUEIREDO, 2010).

Determinada a distribuição empírica, é possível calcular o p -valor comparando o valor de $T = LR(z^*)$ com a distribuição empírica obtida através das simulações, e assim verificar se o *cluster* z^* é significativo.

5 RESULTADOS

Esta seção se destina a apresentar e interpretar os resultados obtidos pelo método de Varredura Circular aplicado ao banco de dados de ocorrências aeronáuticas no Brasil.

5.1 Clusters Significativos

As 2.671 ocorrências estudadas estão distribuídas em 503 municípios do país. Dessas, 589 foram classificadas como acidente aeronáutico e eles estão distribuídos em 346 municípios.

Admitindo um nível de significância de 5%, utilizando o algoritmo de Varredura Circular para encontrar os valores de T e a simulação de Monte Carlo para testar a significância das zonas identificadas, verifica-se que há cinco zonas no país onde a proporção do número de acidentes é significativamente maior do que fora delas.

A zona mais significativa encontrada é a que compreende o círculo de número 1, cujo p -valor é igual a $7,5 \times 10^{-15}$. A segunda mais significativa é a representada pelo número 2 com p -valor igual a $3,2 \times 10^{-7}$. Depois a zona 3 com p -valor dado por $2,5 \times 10^{-5}$, seguida da zona 4, com p -valor igual a $2,6 \times 10^{-5}$. Já a zona cinco, entre as que possuem um p -valor menor que o nível de significância de 5%, foi a menos significativa com um p -valor igual a $3,3 \times 10^{-2}$.

Interessados em obter a lista dos municípios pertencentes a cada um dos clusters podem contatar os autores. As zonas significativas podem ser visualizadas na Figura 3.



FIGURA 3 - Conglomerados de Acidentes Aéreos Significativos no Território Brasileiro.

A Tabela 1 contém o número de casos, c_z , tamanho da população, n_z , proporção de acidentes, p_z , número esperado de acidentes, $\mu_z = p n_z$, o valor da log-verossimilhança, $LLR(z)$, e o p -valor de cada um dos *clusters*.

TABELA 1 - Tabela comparativa dos clusters.

Conglomerado	c_z	n_z	p_z	$p \times n_z$	$LLR(z)$	p -valor
Brasil	588	2665	0,2206	-	-	-
1	124	244	0,5082	53,84	38,1188	7,5x10-15
2	48	76	0,6316	16,77	20,1199	3,2x10-7
3	39	63	0,6190	13,90	15,6918	2,5x10-5
4	23	25	0,9200	5,52	15,6216	2,6x10-5
5	19	29	0,6551	6,40	8,21515	3,3x10-2

Diversos fatores estão associados à maior proporção de acidentes nesses locais. Como podemos observar, o cluster 1, por exemplo, é uma região de baixa densidade populacional, logo, quando acontece uma ocorrência aeronáutica de pequena repercussão, a tripulação pode deixar de notificar o CENIPA desse incidente. Dessa forma, o número de ocorrências fica menor do que o real o que aumenta a proporção de acidentes nesse local. Pode-se supor que no cluster 4, que abrange o interior da Bahia, onde ocorrem menos pousos e decolagens, ocorra o mesmo.

No cluster 2 está a região serrana de Minas Gerais. Nessa área há muita instabilidade meteorológica o que pode influenciar na gravidade das ocorrências.

A área do cluster 3 é marcada por uma intensa atividade agrícola que se constitui em um segmento em que o voo tem maior risco pela proximidade com o solo. Além disso, a região possui grandes variações climáticas em curto espaço de tempo, o que pode representar aumento no risco.

Por fim, no cluster 5, apesar de haver maior fiscalização, está a região de maior volume de tráfego aéreo o que provavelmente causa mais estresse nos profissionais pela complexidade de trabalhar nesta área.

6 CONCLUSÃO

Neste estudo, o algoritmo de Varredura Circular de Kulldorff foi utilizado para a identificação de conglomerados espaciais de acidentes aéreos no Brasil a fim de encontrar áreas em que a proporção de acidentes é significativamente maior do que no resto do país.

Os dados das ocorrências aeronáuticas foram fornecidos pelo CENIPA e, com as informações dos municípios, buscaram-se no IBGE suas coordenadas geográficas. No entanto, perdemos a informação de 43 eventos, pois não foi possível identificar o local exato dessas ocorrências.

O estudo identificou cinco *clusters* significativos mostrando que há regiões no país que precisam de atenção especial. Dessa forma, espera-se que este resultado possa orientar tanto as ações preventivas como a fiscalização de acordo com as características de cada uma dessas áreas.

O cluster número 1 apresentou um *p*-valor extremamente pequeno mostrando que o problema nessa área é ainda mais sensível. Acredita-se que a tripulação não notifica as ocorrências de menor repercussão o que torna a proporção de acidentes neste local maior. Sendo assim, por desconhecer as situações de potencial risco, o CENIPA deixa de analisá-las e, segundo a teoria de Heinrich, deve-se tentar diminuir os incidentes, pois assim, o número de acidentes também tenderia a cair.

As informações deste trabalho são relevantes para a Segurança de Voo e possibilita um melhor direcionamento na adoção de medidas preventivas, além de apontar locais que mais precisam de fiscalização. Logo, espera-se que a prevenção de acidentes aéreos possa ser feita na raiz do problema de forma a reduzir as situações de perigo em vez do dispêndio de esforços em suas consequências.

REFERÊNCIAS

BRASIL. **Código Brasileiro de Aeronáutica**. Lei 7.565, de 19 de dezembro de 1986. Disponível em: <<http://www.cenipa.aer.mil.br/cenipa/index.php/legislacao/category/5-outros>>. Acesso em: 26 nov 2012.

BRASIL. Comando da Aeronáutica. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. **Estatísticas**. Disponível em: <http://www.cenipa.aer.mil.br/cenipa/Anexos/article/18/aviacao_civil.pdf>. Acesso em 20 dez 2012.

BRASIL. **Decreto Nº 87.249**, de 07 de junho de 1982. Disponível em: <<http://www.cenipa.aer.mil.br/cenipa/index.php/legislacao/category/5-outros>>. Acesso em: 26 nov 2012.

BRASIL. Comando da Aeronáutica. **Norma de Sistema do Comando da Aeronáutica 3-1 de 31 de outubro de 2008**. Disponível em: <<http://www.cenipa.aer.mil.br/cenipa/index.php/legislacao/category/1-nscs-norma-do-sistema-do-comando-da-aeronautica>>. Acesso em: 26 nov 2012.

FIGUEIREDO, R. L.. **Detecção de clusters usando a Estatística Scan Espacial Circular em conjunto seletivos e um fator de penalização: a ocupação circular.** 2010. 66 f. Dissertação (Mestrado em Estatística) - Curso de Pós-graduação em Estatística. Universidade Federal de Minas Gerais, Minas Gerais, 2010.

KULLDORF, M. **A spatial scan statistic**, Communications in Statistics – Theory and Methods. v.26, n.6, p.1484-1496. Disponível em: <<http://dx.doi.org/10.1080/03610929708831995>>. Acesso em 22 nov 2012.

WIKIPEDIA. **Herbert William Heinrich**. Disponível em: <http://en.wikipedia.org/wiki/Herbert_William_Heinrich>. Acesso em: 11 de nov de 2012.

IDENTIFICATION OF SPATIAL CLUSTERS OF BRAZILIAN AERONAUTICAL ACCIDENTS

ABSTRACT: This paper aims to verify the existence of spatial clusters of aeronautical accidents in the Brazilian territory, in order to identify the regions in which the proportion of accidents, if one considers the total of occurrences, is significantly bigger than the percentage of the remainder of the country. With such result available, the guidance for the prevention of accidents becomes more accurate, thus strengthening aviation safety in Brazil. For that purpose, this work utilized the Kulldorff's Spatial Scan Statistic, appropriate for the analysis of spatial grouping. The work considered the civil aviation occurrences in the Brazilian territory between the years 2007 and 2011 whose crash-site geographical coordinates were identified.

KEYWORDS: Grouping Analysis. Circular Scan. Spatial Cluster. Flight Safety. Aeronautical Accident.