Risk of Fauna in Brazilian Aviation: Application of Correspondence Analysis to Analyze the Relationship Between Flight Phase and Report Type

Luis Carlos Batista Santos [1]

Cleibson Almeida [2]

Jorge Luiz Farias [3]

Carla Susete Gonçalves Francisco [4]

Beatriz Macedo Coimbra dos Santos [5]

1 Pós-graduado em Segurança de Aviação e Aeronavegabilidade Continuada, Estatístico, Suboficial, Especialista em Controle de Tráfego Aéreo, da Força Aérea Brasileira. E-mail: luiscarloslcbs@fab.mil.br.
2 Doutorando(a) em Matemática Aplicada e Modelação na Universidade Aberta de Portugal, Lisboa.
3 Estatístico, Suboficial, Especialista em Comunicação Aeronáutica, da Força Aérea Brasileira.
4 Phd student, researcher, Quebec city, Canada.
5 Graduanda em Matemática Aplicada na Fundação Getúlio Vargas, Rio de Janeiro.

ABSTRACT: This work presents an application of the statistical technique of Simple Correspondence Analysis to understand the interrelationship of variables and interpret the results analytically for events reports of interest with Fauna in Brazil and the different phases of flight involving the positioning of the aircraft when they are in the same physical space with the fauna. The data used were collected in the Bird Risk Management System (SIGRA), also known as Cenipa 15 (FC15). A total of 13,437 events, dated between January 2011 and December 2016, were considered and R *software* was used for the correspondence analysis. The results showed that there is an interrelationship between event type and flight phase.

KEYWORDS: Fauna Risk, Flight safety, Simple Correspondence Analysis.

1 INTRODUCTION

In the airport area, aircraft and birds compete for physical space at all times and this dispute can cause incidents, serious incidents and aeronautical accidents (MENDONÇA, 2008).

Therefore, in Brasil, the Center for Research and Prevention of Aeronautical Accidents (CENIPA) has a tool for collecting data on events involving aviation and fauna, known as the Cenipa File 15 - FC15 (SIGRA, 2017).

This form is based on an electronic form made available on the internet through the SIGRA (Avian Risk Management System) and the data collected are used to prepare reports for the prevention of aeronautical accidents, as well as the mitigation of the fauna risk in Brazilian aviation. These reports are seen as proactive and predictive forms of prevention within the context of flight safety, as well as, being guides measures for making decisions to avoid future aviation accidents. An example of a report, is the annual Fauna Risk Yearbook, prepared by CENIPA.

According to the Fauna Risk Yearbook 2015, CENIPA file 15 (FC15) is essential for guiding wildlife control measures through information collected through an appropriate form (Oliveira, 2016 and CENIPA, 2017).

Since its implementation in 2011, SIGRA has been collecting information on events of interest to aviation fauna and currently has more than 27,000 records (SIGRA, 2017).

However, it does not make sense to just collect this data. It is necessary to explore in order to try to understand the relationships between variables, to search for trends, to develop facilities in the visualization of information and to make comparisons with already known standards in flight safety. This context is called Data Science (DAVENPORT, 2014, GRUS, 2016).

Considering the above, this work aims to explore from the point of view of data science, through the application of correspondence analysis and the information on aeronautical events collected by the SIGRA System, to reach for a better understanding of the relationship between variables Flight Phase and Report Type. It is expected that a better understanding of this relationship contributes to the improvement of flight safety in the Brazilian aviation.

2 LITERATURE REVISION

2.1 CORRESPONDENCE ANALYSIS

One of the most cited works in the Brazilian literature, on the statistical method of correspondence analysis, was developed by Carvalho and Struchiner (1992). In their research, the authors identified the need to simultaneously study the relationships between a set of variables and for that they applied the correspondence analysis to evaluate the vaccination services of the

National Immunization Program (PNI) in Brasil. After the technical details about the method used, the authors concluded that this type of approach, allows to maximize the understanding, of the relationship between variables and also avoids misinterpretations from empiricism and simpler statistical methods.

Pamplona *et al.* (2007), adopted the correspondence analysis to study the crimes recorded in the metropolitan region of Belém, (Brasi)l, in 2006. The main result obtained by the researchers was a better understanding about the relationship of the variables, for example, the municipality association of "Belém" with the "crimes against the person" and that city of "Ananindeua" had relation with "crimes against the patrimony". Therefore, it is possible to observe that the correspondence analysis served to study the relationship between the variables "local" and "type of crime" at the level of their categories.

In the master's thesis, defended by Cabrita (2012), the theoretical development on correspondence analysis was presented in Portuguese. Therefore, all the details about the contingency table, initial data matrix, masses, row and column profiles, profile cloud, centroid, distance, inertia and the algorithm of the method were presented. In the conclusions, the author emphasizes that the Correspondence Analysis is a useful technique to make available the association between qualitative (categorical) variables.

In more recent works, such as Oliveira (2015), Camelo *et al.* (2016) and Pereira *et al.* (2016), the correspondence analysis was used to analyze the efficiency in public purchases, to classify the wind speed in the Brazilian northeast and to analyze the primary health services offered to alcohol users in Brasil. These three papers show the scope for applying correspondence analysis, as well as being good examples to illustrate the current interest of researchers in different areas of knowledge in the application of this statistical method.

Thus, with the referential presented it is possible to perceive that the use of correspondence analysis is not recent and has been used in several areas in order to understand the relationship between qualitative (categorical) variables.

An important theoretical question on the subject is that the relationship between variables in the correspondence analysis is deeper than in the correlation analysis. While the first one allows to understand the relationship between the different categories of the variables studied, the second allows only to measure the relationship between the variables, without deepening the relations between the items (categories) that make up the variables (HAIR *et al,* 2009).

2.2 AVIATION AND FAUNA

In Brasil, events involving aircraft and wildlife are not rare. Only in 2016 there were 6,023 events, considering collisions, near collisions and sighting of fauna (SIGRA, 2017).

These events are reported from different sources, such as aircraft pilots, air traffic controllers, aerodrome managers, or anyone else who views a wildlife hazard scenario.

According to Li & Li (2010) and Varga *et al.* (2014), events involving aircraft and wildlife can be divided into "birdstrike" and "wildlife strike". While the first refers to events with birds, the second covers events with the other species of fauna.

Although the behavioral fauna factor and the operational factor of the aircraft are almost always present in events of this nature, the main focus of this area of study is the aerodrome infrastructure characteristics (*ABREU ET AL.,* 2017).

According to Abreu *et al.* (2017), the airport areas have the elements that attract the animals to the proximity of the aeronautical operations, such as food, shelter and water.

Therefore, this is the main context on the risk of fauna for aviation and the academic papers dealing with this subject are focused on the following interests:

    a) In the relation between species of fauna involved and place of the aeronautical event (usually aerodromes). For example, Novaes and Alvarez (2014), present the relationship between urban solid waste near an aerodrome with the black-headed vultures in the city of Ilhéus, Brasil.

    b) In the identification of species that affect the aerial operation. For example, in the work of Dove *et al.* Forensic techniques are presented for identification of birds during the accident investigation.

    c) In the search for models for aerodrome risk management. For example, Patrick and Shaw (2012) present a collection of model characteristics for managing aerodrome hazards and analyze their effectiveness. There is also the work of Ntampakis and Biermann (2014), who propose the SMS (Safety Management System) as a suitable model for managing the risk of fauna at aerodromes.

    d) In the development of techniques to mitigate the risk of fauna in aviation. For example, Abreu *et al.* (2017) address risk mitigation through the management of wildlife species that inhabit the airport environment. In order to do so, the authors applied different heights of grass cuttings at an aerodrome with the expectation of discouraging the interest of the birds in that environment and then comparing the effects of each type of cut on the attractiveness of the fauna.

Therefore, it can be seen in the literature consulted that the greatest interest of the academic works in this area are aimed at the identification of species and the mitigation of the fauna risk at aerodromes.

3 METHODOLOGY

From the entire data context in the SIGRA system, this study focuses on two variables: a) type of event and b) flight phase. These variables are categorical, in the condition of nominal qualitative.

The **event type**, or report, is categorized as follows: collision, near collision, and sighting (SIGRA, 2017).

The **flight phase** is characterized as follows: parking, taxi, takeoff, climb, cruise, descent, low altitude navigation, approach and landing (SIGRA, 2017).

During the period of data collection, between 2011 and 2016, different events occurred in 324 Brazilian airfields. Considering that the number of aerodromes with events in this period is excessive and some of them had a low number of events, the Pareto principle was adopted with a focus on reducing the number of aerodromes.

Basically, the Pareto principle considers that 20% causes are responsible for 80% of the effects. Thus, the database was reduced and considered only the aerodromes (20%) more representative as the number of events involving fauna and aircraft (80%).

The application of the Pareto principle allowed to reduce the initial 324 aerodromes to 40. Thus, the 40 aerodromes that remained in the database to be analyzed represent 80% of the events in the period (Figure 1).
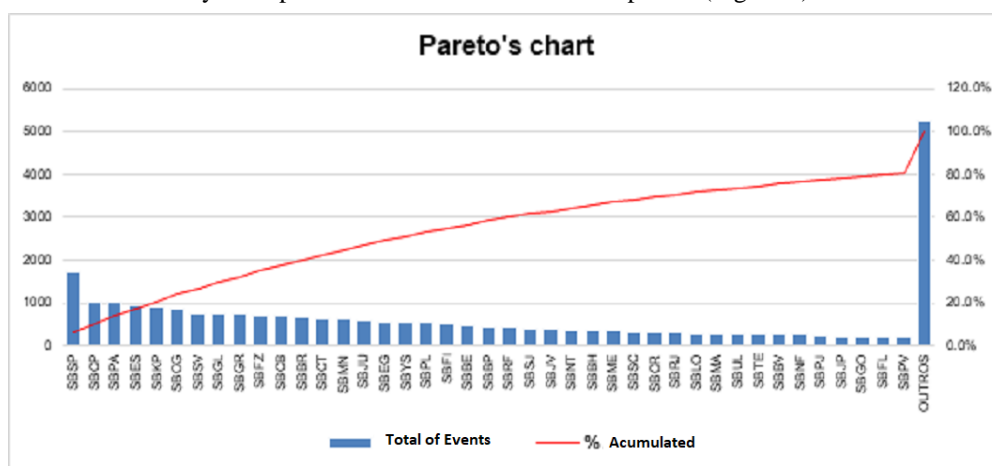
Figure 1: Distribution of events versus Pareto principle.

Table 1, shows tabulated data that was used.

	FLIGHT PHASE	REPORT TYPE			
		COLISION	ALMOST COLISION	SIGHTING	TOTAL
1	ESTA_TAX	114	44	639	797
2	TAKE OFF	2024	539	957	3520
3	CLIMB	182	178	166	526
4	CRUISE	34	133	145	312
5	DESCEND	82	79	172	333
6	NBA	71	119	139	329
7	APROXIMATION	702	1509	1877	4088
8	LAND	2514	365	653	3532
	Total	5723	2966	4748	13437

Table 1: Fauna risk events between 2011 and 2016.

The flight phase "traffic inspection / interflight" was not entered because there was no knowledge of the "place / flight phase" where the event occurred. And in case the flight phase "track review" only occurs in report type "collision and sighting", was also not considered.

3.1   ON THE APPLICATION OF THE METHOD

As the proposal of this work is in the understanding of the interrelationship of type of event with the flight phase and these two variables are nominal qualitative (categorical), a statistical method will be applied for this purpose.

One way to understand this interrelationship is through the multivariate data technique called correspondence analysis (ACS). When one has an interest in the analysis of only two variables, the method is known as simple matching analysis.

According to Hair *et al.* (2009), simple correspondence analysis (ACS) is a multivariate data technique used to verify relationships between nominal data categories in a contingency table.

In this sense, ACS proposes to construct a perceptual map, which is a visual representation of the understandings of objects of an individual in two or more dimensions (Hair et al., 2005).

According to Fávero and Belfiore (2009), the ACS has two fundamental steps: calculation of the measure of association and creation of the perceptual map. To form the basis for association, a chi-square test (χ^2) is done with the objective of standardizing the frequency values. With the standardized measure of association, a distance measure and orthogonal projections are generated for which the categories can be allocated to represent the degree of association reported by the χ^2 distances to be demonstrated by a dimensional space.

According to Fávero and Belfiore (2009), the eigenvalues are found by multiplying matrices, and the square of each eigenvalue is understood as the inertia of the dimensions, which shows importance for each dimension. The division between the inertia of each dimension and the total inertia is the proportion of the variance explained by the dimension. The maximum number of dimensions is estimated by:

- [Minimum (row, column) - 1] = [minimum (8, 3) - 1] = 2

With this result, it is followed for the examination of graphical representation (perceptual map). In this sense, it is necessary to know if there is an association between the variables and the degree of dependence between the variables. For this, the chi-square test is applied.

In this sense, it is necessary to construct a statistical hypothesis test for decision making. The hypotheses tested are:

- H0: the two categorical variables are randomly associated.
- H1: the association between the two categorical variables is not random.

Finally, a perceptual map is created. According to Fávero and Belfiore (2015), the perceptual map serves to visually locate on the graph objects of an individual in two or more dimensions. In the perceptual map each object has a spatial positioning that describes its similarity in relation to other objects, according to the dimensions exposed in the perceptual map.

Statistical software is used to perform all these calculations. In this work, R and Excel were used for tabulation and data analysis, construction of the Pareto chart, application of the method, hypothesis tests and the elaboration of the tables.

4 RESULTS

In this section the results obtained with the application of Correspondence Analysis are presented. The results are presented as follows: a) descriptive analysis and b) correspondence analysis.

4.1 DESCRIPTIVE ANALYSIS

It can be observed in figure 1 that "collision" with fauna occurs more frequently in the "takeoff" and "landing" phases. In fact, in these two phases the aircraft is closer to the ground of the airport area. With this, the aircraft's exposure to the fauna is evidenced, since this is a perfect setting to find food, shelter and water, key items of attraction to birds and land animals.

In the flight phase "approach" there are more "sightings" and "almost collisions" with fauna. In this phase of flight, the pilot of the aircraft fulfills a trajectory or traffic circuit near the aerodrome, which allows an immediate decision to deviate the aircraft before the impact with the animal.

Sightings of birds and land animals are more frequent in the "parking", "taxi", "landing" and "take-off" flight phases. In the courtyard, taxi and airstrip areas are concentrated lawns that attract typical animals and allows the sighting of fauna by the airport team.
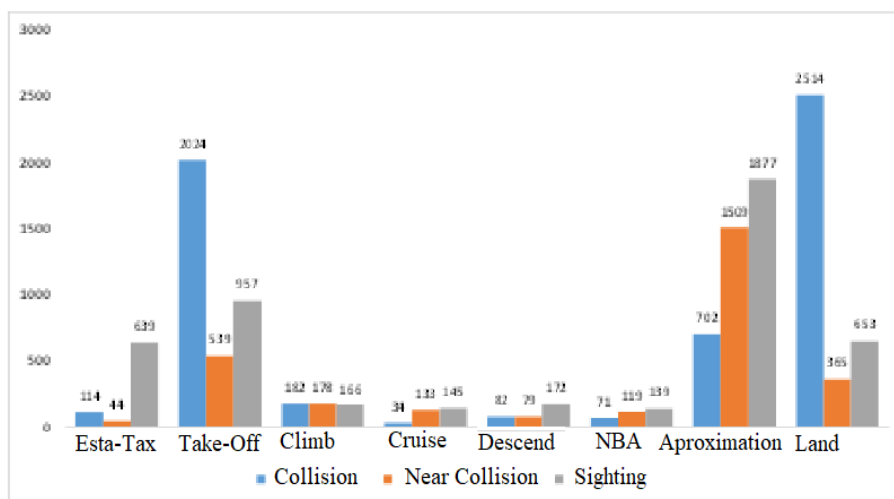
Graph 1: Events per flight phase at the risk of fauna between 2011 and 2016.

Table 2 shows the distribution of the proportions (in percentage) in relation to the totals of each row and the totals of each column respectively (L, C) of the variables. The online analysis shows that there is more "sighting" of fauna (80.2%) than "collision" and "near collision" in the flight and taxi flight phase. The "sighting" and "near collision" with fauna occur in 89.1% of the time, while the collision with birds represents only 10.9% in the "cruise" flight phase.

In the column analysis, it was observed that 50.9% of the almost collisions with birds occur in the flight phase approach and the collision with fauna occurs 79.3% in the events involving the landing and takeoff flight phases. It is also possible to note that 39.5% of sightings occur in the flight approach phase.

Looking at the analysis of the joint distribution of proportions for rows and columns, it is not possible to be sure of the association between the variables.

Report type / Flight phase		Colision	Almost colision	Sighting
Parking and Taxi	Line	14.3%	5.5%	80.2%
	Row	2.0%	1.5%	13.5%
Take-off	Line	57.5%	15.3%	27.2%
	Row	35.4%	18.2%	20.2%
Climb	Line	34.6%	33.8%	31.6%
	Row	3.2%	6.0%	3.5%
Cruise	Line	10.9%	42.6%	46.5%
	Row	0.6%	4.5%	3.1%
Descend	Line	24.6%	23.7%	51.7%
	Row	1.4%	2.7%	3.6%
NBA	Line	21.6%	36.2%	42.2%
	Row	1.2%	4.0%	2.9%
Aproximation	Line	17.2%	36.9%	45.9%
	Row	12.3%	50.9%	39.5%
Land	Line	71.2%	10.3%	18.5%
	Row	43.9%	12.3%	13.8%

Table 2: Joint distribution of ratios for rows and columns.

4.2 CORRESPONDENCE ANALYSIS

The Chi-square test (χ^2) was calculated according to Table 2 and the hypotheses tested were presented in the methodology.

Information	Values
Statistic χ^2	3670.455
Degrees of freedom	14
P-Value	0

Table 3: X² Test.

Thus, the test statistic was significant, presenting p-value <0.05, with 95% confidence, recommending rejection of the null hypothesis. This shows the association between the two categorical variables, which could not be verified only with the information in Table 2.

Thus, standardized residuals adjusted with positive values greater than 1.96, in absolute value, indicate that there is evidence of a significant association between the categories. High residue means greater dependence between the variables, which facilitates the analysis of the perceptual map. Note that values above 1.96 were highlighted in yellow.

In Table 4, it can be seen that the phases of flight takeoff and landing have a strong association with type of collision report. On the other hand, for the flight phases "ascent", "cruise" and "NBA" are associated with near collision, the "approach" has a stronger association with this type of report (near collision). In the "sighting" report type, the variables "parking" and "taxi" presented greater association than the "cruise", "descent" and "NBA" phases. It is also possible to observe that in the "approach" there is a significant association with "near collision".

	FLIGHT FASE	REPORT TYPE		
		COLISION	ALMOST COLISION	SIGHTING
1	ESTA_TAX	-16,65	-11,61	27,3
2	TAKE-OFF	20,82	-11,25	-11,77
3	CLIMB	-3,78	6,63	-1,84
4	CRUISE	-11,45	8,85	4,16
5	DESCEND	-6,71	0,73	6,3
6	NBA	-7,8	6,24	2,56
7	APROXIMATION	-39,4	27,42	16,96
8	LAND	40,01	-19,6	-24,39

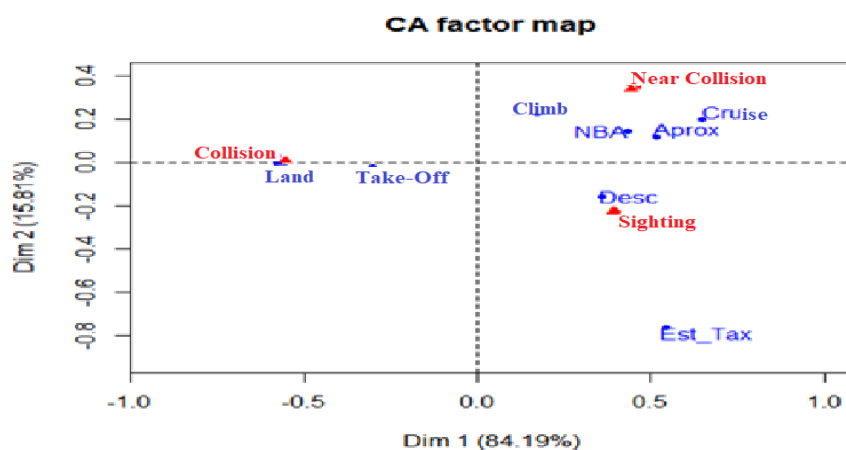**Table 4:** Association between flight phase categories and type of report.

We have below the *output* of the inertial decomposition analysis for the two dimensions, done with software support R with package ca.

Dimenson	Autovalue	%	% Acumulated
1	0.229972	84.2	84.2
2	0.043188	15.8	100.0
Total	0.273160	100.0	

Table 5: Inertial contribution.

As seen in Table 5, the total inertia is 0.273160. The proportions explained in dimensions 1 and 2 correspond respectively to 84.2% (0.229972 / 0.273160), and to 15.8% of total inertia. It is worth remembering that the greater the inertia of the first dimension, the greater the association between the categories arranged in line and in columns.

In the perceptual map (Graph 2), it is graphically emphasized that the collision with fauna happens more frequently in the landing and takeoff, and the near collision occurs in the "climb", "NBA", "approach" and "cruise" phases. The sighting of fauna is more frequent in the descent and especially during the parking and the taxi.

Graph 2: Perceptual map of the association between flight phase and type of report.

5   CONCLUSION

This work presented the application of the statistical method of Correspondence Analysis to identify the relationship between two important variables in the management of fauna risk in Brazilian aviation.

According to a literature review, it was possible to identify the need for studies with a focus on the verification of the behavior of the categorical variables collected by the CENIPA file 05 datasheet. For this purpose, the type variables of the report and flight phase were studied, since they are considered important within of the theme.

In the results we presented exploratory statistical analyzes and the application of the Correspondence Analysis to the data. After all the methodological validation of the presented results, it is possible to corroborate with the flight safety community, as follows:

a) that air carriers should pay attention to:

• In the takeoff and landing procedure, where there are collisions with fauna. Remembering that collision is usually more damaging than other types of reporting;

• In the approach procedure, because this phase of the flight almost collisions occur.

• in the parking lot / taxi of the aircraft, because in this case there is a sighting of fauna inside the airport area and it is an opportunity to identify and mitigate the risk before the animals interfere in flight phases where the event is more harmful.

In this way, the correspondence analysis served to prove the relationship between flight phase and type of report. Understanding this relationship results in valuable information for decision-making by crewmembers and aerodrome managers, serving to disseminate the risk of fauna as it relates to these two variables and may prevent future aeronautical accidents.

Finally, it is proposed the use of this methodology for future studies of the association of other variables that affect the risk of fauna in Brazilian aviation.

REFERENCES

ABREU, T. L. S. et al. Evaluation of different grass height management patterns for bird control in a tropical airport. **Revista Conexão SIPAER**, v. 8, n. 1, p. 68-79, 2017. Disponível em: <http://conexaosipaer.cenipa.gov.br/index.php/sipaer/article/view/363>. Acesso em: 14 jun. 2017.

BRASIL. Comando da Aeronáutica. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. **Plano do Comando da Aeronáutica (PCA) 3-3**: Plano Básico de Gerenciamento de Risco de Fauna. Brasília. 2017. Disponível em: <http://www2.fab.mil.br/cenipa/index.php/legislacao/pca-plano-do-comando-da-aeronautica. Acessado em: 17 jun. 2017.

CABRITA, D. M. D. **Métodos multivariados para variáveis qualitativas**: aplicação ao estudo de variáveis associadas com a avaliação na disciplina de Matemática de uma escola do Ensino Básico no Concelho de Vila Nova de Gaia. Dissertação (mestrado), Mestrado em Estatística, Matemática e Computação, Universidade Aberta de Portugal. 2012.

CAMELO, H. N. et al. Utilização de Análise de Correspondência para Classificação da Velocidade do Vento no Nordeste Brasileiro. **Revista Principia - Divulgação Científica e Tecnológica do IFPB**, [S.l.], n. 31, p. 22-28, dez. 2016. ISSN 2447-9187. Disponível em: <http://periodicos.ifpb.edu.br/index.php/principia/article/view/642>. Acesso em: 27 Nov. 2017.

CARVALHO, M. S.; STRUCHINER, C. J. Análise de Correspondência: Uma aplicação do Método à Avaliação de Serviços de Vacinação. **Cadernos de Saúde Pública**, v. 8, n. 3, p. 287-301, jul/set, Rio de Janeiro. 1992.

CENTRO DE INVESTIGAÇÃO E PREVENÇÃO DE ACIDENTES AERONÁUTICOS [CENIPA]. **Anuário de Risco de Fauna**, 2015. Brasília. 2016. Disponível em: <http://www2.fab.mil.br/cenipa/index.php/estatisticas/risco-da-fauna?download=129:perigo-aviario-e-fauna>. Acesso em: 14 jun. 2017.

CENTRO DE INVESTIGAÇÃO E PREVENÇÃO DE ACIDENTES AERONÁUTICOS [CENIPA]. **Sistema de Gerenciamento de Risco Aviário (SIGRA)**: banco de dados. 2017. Disponível em: <http://sistema.cenipa.aer.mil.br/cenipa/sigra/pesquisa_dadosExt>. Acesso em: 14 jun. 2017.

DAVENPORT, T. H. **Dados Demais**: Como desenvolver habilidades analíticas para resolver problemas complexos, reduzir o risco e decidir melhor. Rio de Janeiro: Campus Elsevier. 2014.

DOVE, C. J.; DAHLAN, N. F.; HEACKER, M. Forensic Bird-strike identification techniques used in an accident investigation at Wiley Post Airport. **Human-Wildlife Conflicts**, v. 3, p. 179-185. 2008.

GRUS, J. **Data Science do Zero**: Primeiras regras com o Python. Rio de Janeiro: Alta Books. 2016.

FÁVERO, L. P.; BELFIORE, P. **Análise de dados**: Técnicas multivariadas exploratórias com SPSS® e Stata®. Rio de Janeiro: Campus Elsevier. 2015.

FÁVERO, L. P.; BELFIORE, P. **Análise de Dados**: Modelagem multivariada para tomada de decisões. Rio de Janeiro: Campus Elsevier. 2009.

HAIR J. J. et al. **Análise multivariada de dados**. 6a Ed. Porto Alegre: Bookman. 2009.

MENDONÇA, F. A. C. SMS for bird hazard: Assessing airlines pilots perceptions. Master of Science in the Department of Aviation. University of Central Missouri. Master Thesis. 2008.

NOVAES, W. G.; ALVAREZ, M. R. D. V. Relação entre o resíduo sólido urbano e urubus de cabeça-preta (Coragyps atratus): um perigo para as aeronaves no aeroporto de Ilhéus (SBIL). **Revista Conexão SIPAER**, v. 5, n. 1, p. 22-29, 2014. Disponível em: <http://conexaosipaer.cenipa.gov.br/index.php/sipaer/article/view/256>. Acesso em: 14 jun. 2017.

NTAMPAKIS, D.; BIERMANN, T. Applying SMS and sustainability principles to airport wildlife hazard management. **Revista Conexão SIPAER**, v. 5, n. 1, p. 8-21, 2014. Disponível em: <http://conexaosipaer.cenipa.gov.br/index.php/sipaer/article/.../255/278>. Acesso em: 14 jun. 2017.

PAMPLONA, V. M. S. et al. Análise de Correspondência dos crimes na região metropolitana de Belém (RMB) no 1º semestre de 2006. In: Simpósio de Pesquisa Operacional da Marinha do Brasil. **Anais...** Rio de Janeiro, 2007.

PATRICK, K.; SHAW, P. Bird strike hazard management programs at airports: what Works? In: 5º Simpósio de Segurança de Voo do Instituto de Pesquisas e Ensaio em Voo. **Anais...** São José dos Campos, São Paulo, Brasil, 2012.