

ANÁLISE DE CLUSTER COMO INSTRUMENTO DE CLASSIFICAÇÃO DE CAVERNAS NO CONTEXTO SOCIOECONÔMICOⁱ

Ricardo José Calembó MARRA - ricardo.marra@icmbio.gov.br, ricardo.marra@uol.com.br, rjcmarra@gmail.com

Doutor em Desenvolvimento Sustentável – CDS/UnB

Instituto Chico Mendes de Conservação da Biodiversidade – ICMBio

Ed. Sede do IBAMA – CECAV – Centro Nacional de Estudo, Proteção e Manejo de Cavernas.

SCEN – Setor de Clubes Esportivo Norte – Trecho 02 - CEP 70.818-900 – Brasília – DF

Tel (61) 3316.1580 - Fax (61) 3316.1370

Abstract

We know that economic growth does not minimize the impacts generated by income disparity and social development in a region exposed. The social issue plays a key role in natural underground cavity. With objective to classify the speleology patrimony on focus of the socioeconomic, 1,169 caves were selected as a sample of the universe of 6,522 natural underground cavities known and registered in Brazil. A questionnaire was developed to assess the major attribute of importance for each cave and identify the perceived relevance (6 levels) of one aspect specified socioeconomic and focus by the existing environmental legislation. The socioeconomic variable was selected to verify the behavior of that variable front the others. Scope of importance (focus) of the cavities (on a local, regional, national, or international scale) was also evaluated. The questionnaire was distributed to eleven specialists, professionals with a wide experience and knowledge in the area of speleology, and they were asked to evaluate each cave in relation to each of researched aspects. The results of these questionnaires were analyzed using the statistical tools of Multivariate Analysis such as cross-tabulated and cluster analysis.

Introdução

O presente artigo tem como objetivo fundamental estabelecer procedimentos científicos para a correta classificação das **cavidades naturais subterrâneas (CNSs)**, utilizando, para isso, critérios de relevância e atributos de qualidade definidos por legislação ambiental, e verificando as características e os perfis desses ativos ambientais frente às questões socioeconômicas.

Esta investigação possibilitou encontrar, mediante uma regra de classificação científica, uma precisa distribuição para CNS. Utilizando o método Análise de Cluster, foi realizada uma modelagem estatística com o elemento amostral cavernas em relação aos grupos de atributos, construindo uma regra de classificação fundamentada na teoria das probabilidades.

Método

Objetivando projetar um sistema de classificação que reflita como as cavidades naturais podem ser distribuídas considerado os aspectos socioeconômicos, utilizou-se a base de dados de Marra (2008). Entretanto com enfoques e aplicações diferentes, haja vista que no primeiro estudo foram utilizadas todas as 16 variáveis alusivas aos critérios de relevância, e para este artigo priorizou-se tão

somente a análise da variável socioeconômica frente às diversas situações.

Esta investigação estabeleceu como ferramenta de classificação a Análise de Cluster (AC) ou conglomerado. Para CORRAR *et. al.* (2007) a análise de conglomerados, ou *clusters analysis*, é uma das técnicas de análise multivariada cujo propósito é reunir objetos, baseando-se nas características dos mesmos. O grupo resultante dessa classificação pode exibir alto grau de homogeneidade interna (*within-cluster*) e alta heterogeneidade externa (*between-cluster*).

A literatura existente admite que a AC e Análise Discriminante (AD) são considerados métodos eficazes de classificação. Entretanto eles deferem entre si.

O método de AC sugere uma classificação ligada a um número conhecido de grupos cujo objetivo operacional é vincular novas observações a cada um desses grupos dadas certas características que os diferenciam.

CORRAR *et. al. op. cit.*, atribui a técnica da AC amoldes mais primitivos, em que nenhuma definição prévia é feita com relação ao número de grupos ou a sua estrutura. *Clustering* (agrupamento) é praticado com base em similaridades ou distâncias (dissimilaridades). Os *inputs* exigidos são medidas

de similaridade ou dados a partir dos quais as similaridades possam ser computadas.

A similaridade entre os objetos é obtida através de coeficientes específicos para cada tipo de variável (discreta, contínua, binária, etc.).

Mingoti (2005) admite que seja necessário considerar medidas que descrevam a similaridade entre elementos amostrais de acordo com as características que neles foram medidas. Se considerarmos que para cada elemento amostral têm-se informações de p -variáveis armazenadas em um vetor, a comparação de diferentes elementos amostrais poderá ser feita através de medidas matemáticas (métricas), que possibilitem a comparação de vetores, como as medidas de distâncias.

Para Mingoti *op. cit.* algumas medidas mais comuns para medir similaridade/dissimilaridade são apresentadas como: Distância Euclidiana (dois elementos amostrais são comparados em cada variável pertencente ao vetor de observações); Distância generalizada ou ponderada (quando se deseja utilizar a ponderação das diferenças das coordenadas dos vetores que estão sendo comparados); Distância de Minkowsky (as distâncias entre os elementos amostrais são armazenadas numa matriz de dimensão chamada de matriz de distâncias); Coeficientes de concordância simples; Coeficiente de concordância positiva; Coeficiente de concordância de Jaccard e Distância Euclidiana média (trata-se de um índice de discordância ou de dissimilaridade). (MINGOTI, 2005, p. 156-164).

Resultados

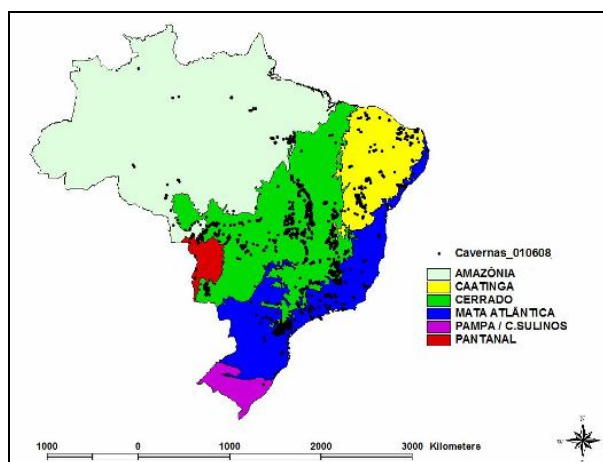
Segundo o Centro Nacional de Estudo, Proteção e Manejo de Cavernas (Cecav, 2008) órgão do Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio)/MMA, são conhecidas no Brasil 6.522 **cavidades naturais subterrâneas (CNS)**, cujos dados estão amplamente disponibilizados a partir da base de dados construída para cadastramento do potencial do País sobre seus ativos ambientais (atualização em 1º/6/2008).

Para Marra (2008) a quantidade conhecida no país não se finda em 6.522 ativos ambientais de CNS, uma vez que o Brasil possui grande potencial espeleológicoⁱⁱ. A distribuição política/regional de CNS no Brasil está apresentada no Quadro 1.

Admite-se que muitas das cavernas conhecidas em cada região ocorre diante da grande

disponibilidade de espeleólogos atuando conjuntamente, evidenciando uma relação direta entre a quantidade de cavernas e esses exploradores.

O Mapa 1 apresenta a distribuição natural do patrimônio espeleológico ao longo de cada um dos biomas. É perceptível notar como o bioma Cerrado recepciona com desenvoltura CNSs (pontos pretos), sendo que ao longo de praticamente toda a extensão dessa grande área geográfica são visualizadas manifestações desses ativos ambientais.



Mapa 1: Distribuição geoespacializada de CNS por biomas brasileiros. Fonte: Marra, 2008, p.98.

Na pesquisa efetuada por Marra *op. cit.*, é possível distinguir que pouco menos da totalidade do patrimônio espeleológico brasileiro (99,95%) referente a 6.519 cavernas estão dispostas em 96,17% do território nacional, distribuídos ao longo dos biomas Cerrado, Mata Atlântica, Caatinga e Amazônia, e de baixa expressão dentro dos biomas Pantanal e Pampa.

Tabela 01: Distribuição do universo de CNS por biomas.

Biomias	Frequência	Porcentagem
CERRADO	4.054	62,16
M. ATLÂNT.	1.043	15,99
CAATINGA	770	11,81
AMAZÔNIA	644	9,87
PANTANAL	08	0,12
C. SULINOS	03	0,05
TOTAL	6.522	100,0

Fonte: (Marra, 2008, p.99, adaptado).

Definição da Amostra

Marra *op. cit.* considera que no cadastro do Cecav *op.cit.* estão registradas um universo de 6.522 cavidades naturais subterrâneas (Base de Dados

Geoespecializados – situação em 1º/6/2008 – data do arquivo original). Crivos de seleção foram empregados considerando aspectos de distribuição por biomas, solos, estados, etc., ficando o universo de cavidades reduzido a um número pouco menor. A problemática dos critérios de relevância foi analisada de acordo com cada situação, remanescendo 1.169 cavernas, número este definido em todos os testes da amostra.

Os critérios de relevância foram extraídos da Resolução Conama nº 347/2004, quando, oportunamente, foram aprovadas e definidas como (*in verbis*):

Art. 2º Para efeito desta Resolução ficam estabelecidas as seguintes definições:

II - cavidade natural subterrânea relevante para fins de anuência pelo Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (Ibama) no processo de licenciamento – aquela que apresente significativos atributos ecológicos, ambientais, cênicos, científicos, culturais ou socioeconômicos, no contexto local ou regional em razão, entre outras, das seguintes características:

- a) dimensão, morfologia ou valores paisagísticos;*
- b) peculiaridades geológicas, geomorfológicas ou mineralógicas;*
- c) vestígios arqueológicos ou paleontológicos;*
- d) recursos hídricos significativos;*
- e) ecossistemas frágeis; espécies endêmicas, raras ou ameaçadas de extinção;*
- f) diversidade biológica; ou*
- g) relevância histórico-cultural ou socioeconômica na região.*

A partir desse ponto foi estabelecido parâmetros no questionário submetido a onze especialista que receberam uma tabela para escolha das relevâncias para cada uma das variáveis/critérios. Definiu-se para cada uma das relevâncias pesos específicos visando posterior mensuração das variáveis, assim estabelecido: **Irrelevante** (Peso 1); **Baixa Relevância** (Peso 3); **Média Relevância** (Peso 5); **Alta Relevância** (Peso 7); **Notável Relevância** (Peso 10) e **Informação Desconhecida** (Peso 0).

A questão socioeconômica como variável estratégica, foi considerada apta para esta pesquisa

diante da importância que este componente exerce junto as CNS, haja vista a grande incerteza que paira sobre como classificar esses ativos ambientais espeleológicos dentro dos atributos estabelecidos na legislação considerando aspectos de mérito tais como são as questões sociais e econômicas.

Na modalagem praticada em Marra *op. cit.* a variável socioeconômica foi definida dentro de um conceito básico objetivando facilitar o julgamento das questões pelos especialistas, o que foi importante para padronizar o entendimento para cada um dos níveis de relevância, assim definido:

Irrelevante: *Cavidade inserida num contexto socioeconômico onde não se verifica demandas de uso e é irrisória a pressão para exploração, seja do ambiente propriamente dito, seja dos recursos naturais em sua área de influência.*

Baixa relevância: *Cavidade inserida num contexto socioeconômico onde se sabe que pelo menos uma vez ao ano (feriados prolongados) há pressão turística para exploração do ambiente propriamente dito e/ou dos recursos naturais em sua área de influência. Eventualmente há pressão para exploração mineral.*

Média relevância: *Cavidade inserida num contexto socioeconômico onde se sabe que nos feriados prolongados e eventualmente nos finais de semana há pressão turística para exploração do ambiente propriamente dito e/ou dos recursos naturais em sua área de influência. A pressão de exploração mineral se processa nos últimos 5 anos.*

Alta relevância: *Cavidade inserida num contexto socioeconômico onde se sabe que nos feriados prolongados e na maioria dos finais de semana há pressão turística para exploração do ambiente propriamente dito e/ou os recursos naturais em sua área de influência. A pressão de exploração mineral se processa nos últimos 10 anos.*

Notável relevância: *Cavidade inserida num contexto socioeconômico onde se sabe que nos feriados prolongados e todo final de semana há pressão turística para exploração do ambiente propriamente dito e/ou os recursos naturais em sua área de influência. A pressão de exploração mineral se processa nos últimos 20 anos. (MARRA, 2008, p. 184).*

No mesmo questionário, foram ainda agregadas questões visando determinar os atributos de qualidade para a cavidade trabalhada. Os atributos

foram também estabelecidos na Resolução Conama nº 347/2004 tais como: ecológicos, ambientais, cênicos, científicos, culturais ou socioeconômicos, conforme descrito no art. 2º, inciso II, assim descrito (*in verbis*):

Art. 2º Para efeito desta resolução ficam estabelecidas as seguintes definições:

II – [...] cavidade natural subterrânea relevante para fins de anuência ...no processo de licenciamento - aquela que apresente significativos atributos ecológicos, ambientais, cênicos, científicos, culturais ou socioeconômicos, no contexto local ou regional [...] (grifo nosso).

A padronização dos conceitos definidos em Marra *op. cit.* para os atributos visando subsidiar a decisão dos especialistas também foi algo importante a solucionar, ficando assim estabelecido:

Atributo Ecológico: *Quando determinada cavidade natural subterrânea contribui para a manutenção do equilíbrio ecológico de sua área de influência e vice-versa, bem como colabora nas interações entre a fauna, flora, demais recursos abióticos da região, além de outros atributos naturais existentes.*

Atributo Ambiental: *Quando determinada cavidade natural subterrânea está encravada em área merecedora de ser consagrada a criação de unidade de conservação, ou já está inserida nos limites uma categoria de UC, objetivando proteger e manter a diversidade biológica, bem como os recursos naturais e culturais associados, sendo essencial ser manejada por meios jurídicos ou outros eficazes da política ambiental brasileira.*

Atributo Cênico: *Quando determinada cavidade natural subterrânea reúne atributos especiais de beleza cênica em seu interior, tais como espeleotemas raros e/ou exuberantes, lagos translúcidos, cachoeiras, galerias ornamentadas e/ou outros atributos de embelezamento no seu interior.*

Atributo Científico: *Quando determinada cavidade natural subterrânea reúne atributos especiais para o desenvolvimento da ciência, seja para produzir e aplicar em objetivos práticos ou para a produção de informações científicas rigorosas das ciências da natureza, bem como dentro do campo de interesse de outras ciências afins, ou ainda servir como laboratório de pesquisa.*

Atributo Histórico-Cultural: *Quando determinada cavidade natural subterrânea é parte integrante da história regional e de um modo geral contém valores simbólicos imateriais significativos, tais como expressões, conhecimentos, práticas religiosas, além dos valores materiais do tipo bens culturais e outros atributos de interesse histórico-cultural.*

Atributo Socioeconômico: *Quando determinada cavidade natural subterrânea é utilizada com finalidade socioeconômica e/ou pode contribuir para trazer subsistência às comunidades do entorno, ou ainda está inserida em área de interesse da mineração e/ou outros projetos de benefício social ou econômico. (MARRA, 2008, p. 186).*

A preocupação para padronização desses conceitos junto aos especialistas, notadamente no contexto socioeconômico, se realizou considerando o uso atual das CNS, embora em muitas situações o habitual uso da cavidade pode servir para definição de variáveis cujas derivadas parciais são parâmetros ou funções importantes para constatações do uso potencial.

Análise dos resultados

Processando as análises nesta pesquisa, notadamente a respeito da distribuição da graduação de relevâncias das CNS amostradas no contexto da variável Socioeconômica por Estado (Tabela 2), verifica-se que Minas Gerais confirmou sua hegemonia em abrigar 411 CNS (35,16%), seguidos de São Paulo com 271 (23,18%) e Rio Grande do Norte com 231 (19,76%) dentre os três principais Estados amostrados respectivamente em ocorrência de cavernas no âmbito deste subconjunto.

Isso significa dizer, em que pese manifestar 615 cavidades (52,61%) na classificação de irrelevantes no contexto estudado, importa consignar que 252 cavernas (21,56%) deste total ficaram eleitas dentro dos padrões de relevância (notável, alta e média) para situações de utilização com finalidade socioeconômica.

Tabela 2: Distribuição da gradação de relevâncias de CNS amostradas em função da Var. Socioeconômica por Estado.

Est	NR	AR	MR	BR	IR	ID	Total
MG	11	18	24	32	275	51	411
SP	22	15	74	130	30	0	271
RN	10	0	7	5	209	0	231
GO	0	4	2	18	44	1	69
MT	5	3	7	12	20	3	50
MS	1	4	10	24	4	0	43
DF	23	1	3	7	4	0	38
BA	0	1	2	10	9	0	22
ES	2	1	0	4	8	0	15
CE	1	0	1	2	4	0	8
RJ	0	0	0	1	3	0	4
AL	0	0	0	2	1	0	3
SE	0	0	0	0	3	0	3
TO	0	0	0	0	1	0	1
Tot	75	47	130	247	615	55	1.169

ID = Informação desconhecida, IR = Irrelevante,
BR = Baixa relevância, MR = Média relevância,
AR = Alta relevância, NR = Notável relevância.

Fonte: Dados da pesquisa.

A Tabela 3 apresenta a distribuição amostral de CNS por biomas. No exame dessa ordenação distingue-se que o bioma cerrado com 509 cavidades amostradas detém parcela significativa (43,54%) das ocorrências, seguido dos biomas Mata Atlântica (364; 31,14%) e Caatinga (252; 21,52%) respectivamente. Ocorre que se aplicado um recorte estatístico (análise contingente) com as relevâncias de CNS no enfoque socioeconômico, então é possível obtermos notáveis relevâncias majoritariamente relacionadas ao cerrado (36; 3,08%), confirmando a supremacia. Destaca-se que 238 cavernas (20,36%) deste total ficaram eleitas dentro dos padrões de relevância (notável, alta e média) quando destacados os interesses relacionados com finalidade socioeconômica, notadamente aos três biomas de maior expressão em ocorrências, dentre os quais se destaca Cerrado, Mata Atlântica e Caatinga.

Constatações como essa reforçam a necessidade de se atentar para uma gestão ambiental mais expressiva em 20% de cavernas com índices probabilísticos importantes de relevância e que obviamente estão apontando a existência de um potencial de uso socioeconômico previsto para as cavernas notadamente dentro dos biomas Cerrado, Mata Atlântica e Caatinga.

Tabela 3: Distribuição da gradação de relevâncias das CNS amostradas na reunião da Var. Socioeconômica com Biomas.

Relev.	Cerr	M.Atl	Caat	Pant	Tr-CC	Total
NR	36	24	12	1	2	75
AR	15	23	1	1	7	47
MR	36	84	7	0	3	130
BR	76	139	10	19	3	247
IR	291	94	222	4	4	615
ID	55	0	0	0	0	55
Total	509	364	252	25	19	1.169

Cerr=Bioma Cerrado. M.Atl=Bioma Mata Atlântica.

Caat=Bioma Caatinga. Pant=Bioma Pantanal.

Tr-CC=Bioma Transição Cerrado/Caatinga.

ID = Informação desconhecida, IR = Irrelevante,

BR = Baixa relevância, MR = Média relevância,

AR = Alta relevância, NR = Notável relevância.

Fonte: Dados da pesquisa.

Com o objetivo de agrupar o conjunto de variáveis definidas no contexto dos critérios de relevância (Seção 2.1) utilizou-se a análise de clusters não-hierárquicos (*K-means*) com o software SPSS 15.0, sendo selecionada a opção “iteração e classificação” e limitado o número de iterações a 10. A técnica de partição *K-means* consiste na iteração progressiva dos casos, iniciando-se de maneira arbitrária, até que se alcance a estabilidade, com um número predeterminado de classes, sendo a variância intragrupos mínima.

No presente trabalho utilizou-se os centróides dos clusters ótimos, predeterminando o número de 03 (três) clusters para execução da modelagem não-hierárquica.

A classificação de CNS utilizando o método de AC possibilitou identificar perfis que melhor entre si assemelham ou não (similaridade/dissimilaridade) para formarem grupos (agrupamentos) objetivando identificar o posicionamento correto desses ativos ambientais para corretas classificações frente às diversas situações.

O passo inicial da investigação em AC é conhecer se há sobreposições entre os atributos definidos em cada uma das variáveis examinadas objetivando assim, avaliar dentro de um intervalo de confiança aceitável o relacionamento das médias calculadas pela modelagem.

A Figura 1 apresenta o diagrama com os intervalos de confiança obtido nesta pesquisa. Nota-se que o intervalo que menos se sobrepõe refere-se exatamente ao atributo socioeconômico (em destaque) quando avaliado sob luz da variável socioeconômica. Este teste possibilitou inferir com 95% de confiança que a média relacionada com o atributo socioeconômico se sobrepõe apenas com o

atributo cênico no caso das relevâncias de valores medianos (Peso 5), ficando isolado para bem classificar cavernas com perfil socioeconômico nas categorias de relevâncias notável e alta.

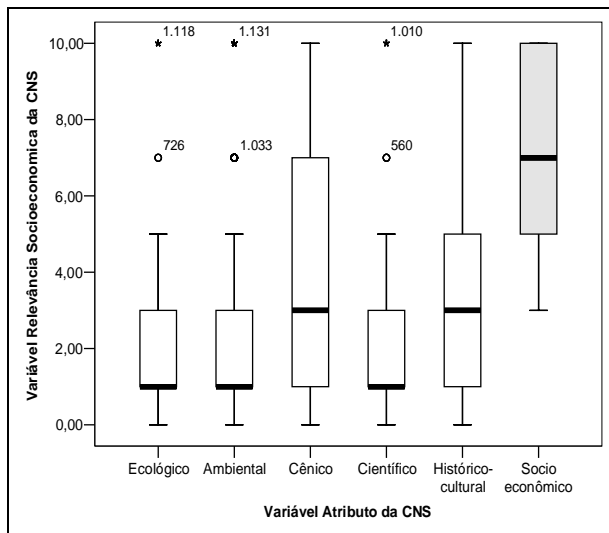


Figura 1: Diagrama para Intervalos de Confiança das CNS amostradas considerando a Var. socioeconômica e Atributos de qualidade. Fonte: Dados da pesquisa.

Outro importante teste executado na AC, refere-se ao diagrama de barras de erro. A Figura 2 apresenta um diagrama de barras de erro demonstrando que existem barras distintas para cada relevância da variável socioeconômica se comparado com os atributos definidos na pesquisa. Essa ilustração permitiu demonstrar que o diagrama de barras de erro com delineamento entre os atributos de CNS possui intervalos das médias de confiança em sobreposição com a maioria dos atributos, notadamente o atributo histórico-cultural com os demais, e bastante distinto se confrontado com o atributo socioeconômico, o que nos permite inferir que não estamos restritos às médias comuns quando se trabalha com intervalos de confiança para questões sociais e econômicas que envolvem cavernas, evidenciando distinções importantes.

A Tabela 4 mostra o resultado da AC para CNSs amostradas com variável socioeconômica agrupadas por Estados. A partir daqui os testes de agrupamento de cavernas por AC exibiram maior distribuição no Cluster 2. Este cluster agrupou 670 cavernas (57,31%) do total amostrado, seguido do Cluster 3 com 424 cavidades (36,27%) e o Cluster 1 com 75 (6,42%).

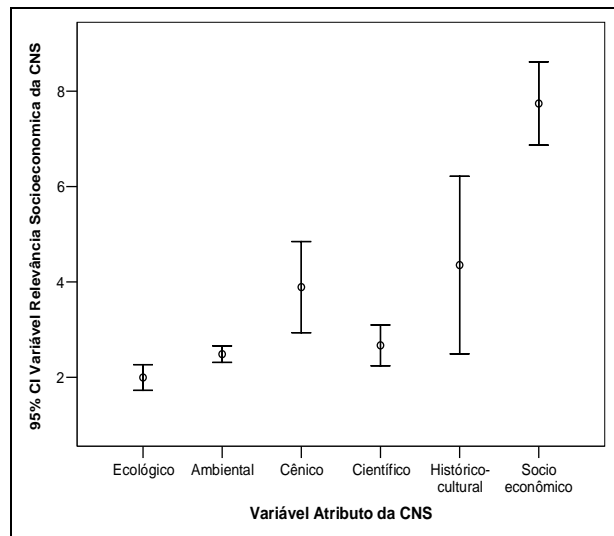


Figura 2: Diagrama de barras de erro das CNS amostradas considerando a Var. socioeconômica e Atributos de qualidade. Fonte: Dados da pesquisa.

Nele a modelagem identificou os Estados de Minas Gerais, Rio Grande do Norte e Goiás como aqueles que foram apontados como de interesse social e econômico relacionados com cavernas. Por outro lado também pode-se perceber que o Cluster 2 apresenta alguns Estados com certa equanimidade, nos sugerindo avaliar que neste agrupamento configura-se certa conformidade de ações identificadas nas relevâncias socioeconômicas de cavernas.

Tabela 4: Resultado da Análise de Cluster para CNS amostradas com Var. Socioeconômica e Estado.

ESTADO	CLUS1	CLUS2	CLUS3	TOTAL
MG	11	326	74	411
SP	22	30	219	271
RN	10	209	12	231
GO	-	45	24	69
MT	05	23	22	50
MS	01	-	38	39
DF	23	04	11	38
BA	-	09	13	22
ES	02	08	05	15
CE	01	04	03	08
RJ	-	03	01	04
AL	-	01	02	03
SE	-	03	-	03
TO	-	01	-	01
TOTAL	75	670	424	1.169

Fonte: Dados da pesquisa.

A Tabela 5 exhibe o resultado da AC para CNSs amostradas com variável socioeconômica agrupadas por biomas. Conforme discutido anteriormente na exploração dos dados, a distribuição amostral de CNS por biomas demonstrou uma frequência clara com distinção ao bioma cerrado (509 cavidades)

representando (43,54%) das ocorrências. Ocorre que o resultado obtido no Cluster 2 sugere uma inversão deste cenário, ou seja passa o bioma Caatinga a caracterizar uma maior expressão de atividades socioeconômicas relacionadas com cavernas, se comparado p.ex. com o da Mata Atlântica, teoricamente um bioma com projeções antrópicas mais patente.

Tabela 5: Resultado da Análise de Cluster para CNS amostradas com Var. Socioeconômica e Biomas.

BIOMAS	CLUS1	CLUS2	CLUS3	TOTAL
CERRADO	36	346	127	509
M.ATL.	24	94	246	364
CAATINGA	12	222	18	252
TRNS-CC	02	04	13	19
PANTANAL	01	04	20	25
TOTAL	75	670	424	1.169

A Tabela 6 apresenta o resultado da AC para CNSs amostradas com variável socioeconômica e Atributos de qualidade, e a Tabela 7 o resultado validado desta AC. Os efeitos da modelagem sobre estes aspectos demonstraram que os atributos ambiental, ecológico e científico (Cluster 2) foram os que mais se expressaram em cavernas sob enfoque da variável socioeconômica. Um resultado não usual, já que se presume cavidades com potencial de uso inseridas em contextos de atributos afins. Ocorre que os valores imputados nos critérios de relevância para a variável foram significativos nos atributos, somado ao fato da grande frequência de cavidades dentro dos mesmos, o que forçou naturalmente a classificação desses ativos em consonância com os centróides de cada uma das distâncias medidas.

Tabela 6: Resultado da Análise de Cluster para CNS amostradas com Var. Socioeconômica e Atributos.

ATRIBU-TOS	CLUS1	CLUS2	CLUS3	TOTAL
AMB.	31	416	264	711
ECOL.	08	160	66	234
CÊNICO	10	27	18	55
CIENT.	07	61	53	121
HC	04	06	07	17
SOCIOEC	15	00	16	31
TOTAL	75	670	424	1.169

Fonte: Dados da pesquisa.

Tabela 7: Resultado Validado da Análise de Cluster para CNS amostradas com Var. Socioeconômica e Atributos.

ATRIBU-TOS	CLUS1	CLUS2	CLUS3	TOTAL
AMB.	18	570	359	947
ECOL.	00	67	24	91
CÊNICO	02	00	02	04
CIENT.	00	19	12	31
HC	09	14	16	39
SOCIOEC	46	00	11	57
TOTAL	75	670	424	1.169

Fonte: Dados da pesquisa.

Conclusão

A análise de cluster mostrou-se um procedimento eficaz para classificação de cavernas demonstrando ser uma ferramenta da estatística de ótima confiabilidade e robustez. Aplicações dessa metodologia permitiram concluir que, ao se considerar apenas a componente socioeconômica, podemos destacar importantes cenários para as cavernas no exame desta variável, destacando principalmente:

1. Se forem considerados os dez primeiros entes da Federação (MG, GO, TO, PA, SP, BA, RN, MT, PR e MS) é possível abranger 97% das ocorrências de CNSs conhecidas do Brasil;
2. O bioma Cerrado abriga 4.054 CNSs (62,16%) do total do País (Tabela 01), o que faz desse bioma o de maior importância para o patrimônio espeleológico nacional;
3. Cavernas agrupadas no Cluster 2 (Tabela 5) aponta o cerrado e caatinga como biomas estratégicos na aplicação de uma gestão ambiental mais eficaz, subsidiando tomadas de decisão importantes;
4. 238 cavernas identificadas nos padrões de relevância (notável, alta e média) foram classificadas em três biomas de maior expressão em ocorrências, dentre os quais destaca-se Cerrado, Mata Atlântica e Caatinga, determinando a necessidade de se estabelecer uma gestão ambiental mais expressiva, já que 20% dessas cavernas estão sob alvo socioeconômico (Tabela 3);
5. A classificação de 252 cavernas eleitas dentro dos padrões de relevância (notável, alta e média) para situações de utilização com finalidade socioeconômica, apontam para um índice de 22% de ocorrências em área de utilidade para a mineração, o que pode significar interesses

- sociais e/ou econômicos para as comunidades de entorno (Tabela 2);
6. Ocorreu um melhor agrupamento de cavernas nos Estados de Minas Gerais, Rio Grande do Norte e Goiás como aqueles que foram apontados como de interesse social e econômico relacionados com cavernas (Tabela 4);
 7. As cavernas ficaram bem agrupadas (Cluster 2) dentro dos atributos ambiental, ecológico e científico quando avaliadas no contexto da variável socioeconômica (Tabelas 6 e 7); e
 8. O Cluster 2 exibiu maior distribuição de CNS, agrupando 57,31% do total amostrado, seguido do Cluster 3 com 36,27% e o Cluster 1 com 6,42%.

Referências

- CECAV (Centro Nacional de Estudo, Proteção e Manejo de Cavernas). Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio). **Relatório demonstrativo da situação atual das cavidades naturais subterrâneas por unidade da federação. Estado da Bahia.** 2008. Disponível em: <http://www.icmbio.gov.br/cecav>.
- CORRAR, L. J.; PAULO, E.; FILHO, J. M. D. **Análise multivariada – para os cursos de administração, Ciências Contábeis e Economia.** São Paulo: Fundação Instituto de Pesquisas Contábeis, Atuárias e Financeiras (FIPECAFI); Ed. Atlas, 2007.
- JOHNSON, R. A., WICHERN, D. W. Applied multivariate statistical analysis. Engle-wood Cliffs: Prentice-Hall, 1982. In: FREI, F. **Introdução à análise de agrupamentos – Teoria e prática.** São Paulo: Ed. Unesp, 2006.
- MARRA, R. J. C. **Critérios de relevância para classificação de cavernas no Brasil.** Tese de doutorado. Centro de Desenvolvimento Sustentável. Universidade de Brasília, 2008. 393 p. il. Brasília.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada. Uma abordagem aplicada.** Belo Horizonte: Ed. UFMG, 2005.
- REIS, E. **Estatística Multivariada Aplicada.** 2.ed. Lisboa: Ed. Silabo Ltda, 2001.

ⁱ Utilizou-se para produção deste artigo a base de dados trabalhada no software SPSS – *Statistical Package for the Social Sciences* e empregada para elaboração da tese de doutorado do autor intitulada *Critérios de Relevância para Classificação de Cavernas no Brasil* - Universidade de Brasília (UnB) CDS – Centro de Desenvolvimento Sustentável (Defesa em 15/12/2008).

ⁱⁱ Qualquer cálculo do número probabilístico de cavernas estimado no país que aponte para o potencial espeleológico, deve ser realizado considerando regiões com mesma litologia e características ambientais, aplicando-se preferencialmente modelos estatísticos que apresentem índices de confiabilidade, sob o risco de se produzir erros baseado em inferências empíricas que não refletem a realidade.