

AUTOMAÇÃO DO RANQUEAMENTO QUALITATIVO DE ÁREAS ESPECIALMENTE PROTEGIDAS DO ESTADO DO PARANÁ ATRAVÉS DA ANÁLISE FATORIAL

Furtado, E. M.
PpgMNE/UFPR-UnicenP
efurtado@bsi.com.br

Chaves Neto, A.
DEST/UFPR
anselmo@est.ufpr.br

Domingues, Z.H.
PpgMAD/UFPR

Hosokawa, R. T.
PpgEngFlo/UFPR

R. Des. Motta, 2219/ap.704 - CEP 80420-190 Curitiba (PR) Brasil

ABSTRACT

The municipal districts of Paraná that possess in its territories Especially Protected Areas, they have, through the Complemental Law N.º 59/91, known as Law of Ecological ICMS, the right to a percentile of the resources of the Tax on Circulation of Goods and Services - ICMS. The evaluation of each area is made by the Environmental Institute of Paraná - IAP, that accomplishes an evaluation of the quality annually, determining each one of them a score. This score determines the financial reviews to the municipal district. Therefore, it is proposed in the present work a multivariate statistical technique, denominated Factor Analysis, that, through a computacional automatic program, accomplish the qualitative ranking of areas of Faxinais in the state. Faxinais constitutes a special type of protected area. They have been analyzed 80 variables of 20 observed Faxinais, referring to agriculture year from August 1997 through July 1998. Using the Factor Analysis, the reduction of the 80 variables was gotten for just 17 factors. The considered factors contributed to an explanation of 97,7649% of the total of the sample variance.

I. INTRODUÇÃO

Segundo a Federação brasileira, cada estado pode definir um perfil de distribuição dos recursos do ICMS – Imposto sobre Circulação de Mercadorias e Serviços – a que os municípios têm direito. O Estado do Paraná, pioneiramente, definiu pela Lei Complementar N.º 59/91, conhecida como Lei do ICMS Ecológico, que 5% dos recursos devem ser distribuídos segundo um critério ambiental. Por essa lei, metade dos 5% são distribuídos aos municípios que possuem em seus territórios mananciais, e a outra metade, aos municípios que possuem Áreas Especialmente Protegidas, caracterizadas pelas Unidades de Conservação. O Instituto Ambiental do Paraná – IAP realiza anualmente a avaliação do nível de qualidade de conservação das Áreas Especialmente Protegidas. A avaliação possibilita um repasse financeiro ao município.

O objetivo deste trabalho foi a partir das variáveis sugeridas pelo IAP, em grande parte qualitativas, quantificá-las e através de escores estabelecer um *ranking* classificatório de todas as áreas avaliadas. Além disto, o procedimento de cálculo foi automatizado por meio de um programa computacional a ser utilizado pelo IAP. Na solução para o problema de classificação das Áreas Especialmente Protegidas foi utilizada a Análise Fatorial, técnica estatística multivariada. Este trabalho descreve o procedimento estatístico na seção II, bem como a coleta e compilação dos dados. Na seção III são apresentados os resultados obtidos.

II. MATERIAL E MÉTODOS

2.1) Histórico

Até a metade do século XX, a aplicação de qualquer método iterativo demandava um enorme esforço de tempo e matemáticos para ser completado. Muito da Ciência conhecida não era utilizado na prática devido às dificuldades na computação dos resultados. O esforço anglo-americano na busca de desenvolvimento no período entre a Primeira e Segunda Grandes Guerras Mundiais, propiciaram a construção de uma

máquina que veio alterar muito da vida humana. Essa máquina é hoje conhecida como computador eletrônico.

Do primeiro computador a válvulas até os mais modernos computadores atuais, muito progresso científico foi feito. Com o avanço tecnológico, as técnicas que na época eram consideradas inviáveis pela grande quantidade de cálculos passaram a ser mais utilizadas. A velocidade de processamento foi crescendo, e com ela, os custos operacionais caíam para valores bem menores. Algumas técnicas estatísticas baseadas em métodos recursivos aguardaram muitas décadas até que pudessem se desenvolver plenamente. Dentre estas técnicas, uma que encontrava-se completamente construída teoricamente por volta da década de 30, mas só teve aplicação plena no final dos anos 50 com a computação eletrônica, foi a Análise Fatorial.

A Análise Fatorial é uma técnica estatística multivariada, que trata do relacionamento entre conjuntos de variáveis. Inicialmente foi desenvolvida por Spearman com trabalhos intitulados “*General intelligence objectively determined and measured*”, de 1904 e “*The Abilities of Man*” de 1926. Mais tarde teve contribuições de outros psicólogos tais como Thomson, Thurstone e Burt, que a desenvolveram como uma técnica pioneira, tratando de hipóteses sobre a organização da habilidade mental, sugerida pelo exame das matrizes de correlação ou covariância para conjuntos de testes de variáveis cognitivas.

A essencial proposta da Análise Fatorial é descrever, se possível, a estrutura de covariância do relacionamento entre várias variáveis em termos de um número menor de variáveis não observáveis, denominadas fatores. Com esses fatores, procura explicar a estrutura de covariância das variáveis observadas.

2.2) Modelo Fatorial Ortogonal

Sendo o vetor aleatório observável $\tilde{X}$, com p componentes, média $\tilde{\mu}$ e covariância $\tilde{\Sigma}$, o Modelo Fatorial postula que $\tilde{X}$ é linearmente dependente de algumas variáveis aleatórias não observáveis $F_1, F_2, \dots, F_m$, chamadas fatores comuns, e p fontes de variação $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas erros ou fatores específicos. Em particular, o Modelo de Análise Fatorial é

$$\underset{(p \times 1)}{\tilde{X}} = \underset{(p \times 1)}{\tilde{\mu}} + \underset{(p \times m)}{\tilde{L}} \cdot \underset{(m \times 1)}{\tilde{F}} + \underset{(p \times 1)}{\tilde{\varepsilon}}$$

com $m < p$ e, ainda μ_i é a média da variável i ; $i = 1, 2, 3, \dots, p$, ε_i é o i -ésimo fator específico, F_j é o j -ésimo fator comum; $j = 1, 2, 3, \dots, m$, e λ_{ij} é o carregamento da i -ésima variável no j -ésimo fator. Os não observáveis vetores aleatórios $\tilde{F}$ e $\tilde{\varepsilon}$ satisfazem as propriedades:

- $\tilde{F}$ e $\tilde{\varepsilon}$ são independentes;
- $E(\tilde{F}) = \tilde{0}$, $Cov(\tilde{F}) = I$;
- $E(\tilde{\varepsilon}) = \tilde{0}$, $Cov(\tilde{\varepsilon}) = \Psi$, onde Ψ é uma matriz diagonal com a variância específica ψ_i na diagonal principal.

Com as suposições do modelo, tem-se $\tilde{\Sigma} = Cov(\tilde{X}) = \tilde{L}\tilde{L}' + \Psi$ e $Cov(\tilde{X}, \tilde{F}) = \tilde{L}$.

A porção da variância de uma variável X_i , advinda como contribuição dos m fatores comuns, é denominada comunalidade i , e denotada por h_i^2 . A porção da variância desta variável atribuída ao fator específico ε_i é chamada de imparidade ou, mais

frequentemente, variância específica que é denotada por ψ_i . Em particular, para a variável i , tem-se $Var(X_i) = (\text{comunalidade } i) + (\text{variância específica } i)$. Logo, $\sigma_i^2 = \sigma_{ii} = h_i^2 + \psi_i$, $i = 1, 2, \dots, p$.

2.3) Número de Fatores Comuns

O número de fatores adotado é um quesito extremamente importante na análise. No presente estudo, dentre os vários critérios que poderiam ter sido adotados, o critério considerado foi o de que o número de fatores m é igual ao número de autovalores maiores que a média dos autovalores. Uma vez que a análise foi efetuada a partir da matriz de

correlação, a média dos autovalores para R é $\frac{\sum_{i=1}^p \lambda_i}{p} = 1$. Assim foram considerados apenas os autovalores maiores que 1. Este critério foi proposto a primeira vez por Guttman em 1954. Kayser, em trabalhos datados de 1960, apresentou boas razões para se confirmar a confiabilidade do critério, mostrando que funciona muito bem na prática.

2.4) Métodos de Estimação

O Modelo Fatorial Ortogonal procura representar adequadamente os dados com um pequeno número de fatores não observáveis. A matriz de covariância S é um estimador da matriz populacional Σ desconhecida. Se Σ desvia significativamente de uma matriz diagonal, então o Modelo Fatorial pode ser utilizado, e o problema inicial é uma estimação dos carregamentos fatoriais λ_{ij} e variâncias específicas ψ_i . O método adotado na estimação dos parâmetros foi o Método de Componentes Principais.

A Análise de Componentes Principais de uma matriz de covariância amostral S é realizada em função de seus pares de autovalor-autovetor $(\hat{\lambda}_1, \hat{e}_{\sim 1}), (\hat{\lambda}_2, \hat{e}_{\sim 2}), \dots, (\hat{\lambda}_p, \hat{e}_{\sim p})$,

onde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$. Sendo m o número de fatores comuns, a matriz dos

carregamentos fatoriais estimados $\{\hat{\lambda}_{ij}\}$ é dada por $\hat{L} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \cdot \hat{e}_{\sim 1} & \sqrt{\hat{\lambda}_2} \cdot \hat{e}_{\sim 2} & \dots & \sqrt{\hat{\lambda}_m} \cdot \hat{e}_{\sim m} \end{bmatrix}$.

As variâncias específicas estimadas são produzidas pelos elementos diagonais da matriz

$$S - \hat{L} \hat{L}', \text{ então } \hat{\Psi} = \begin{bmatrix} \hat{\psi}_1 & 0 & \Lambda & 0 \\ 0 & \hat{\psi}_2 & \Lambda & 0 \\ M & M & O & M \\ 0 & 0 & \Lambda & \hat{\psi}_p \end{bmatrix} \text{ com } \hat{\psi}_i = s_{ii} - \sum_{j=1}^m \hat{\lambda}_{ij}^2.$$

As comunalidades são estimadas através de $\hat{h}_i^2 = \hat{\lambda}_{i1}^2 + \hat{\lambda}_{i2}^2 + \dots + \hat{\lambda}_{im}^2$.

2.5) Rotação de Fatores

Os carregamentos fatoriais estão sujeitos a uma rotação de fatores sem alteração das comunalidades e variâncias específicas. O objetivo principal é uma melhor interpretação e, conseqüentemente, uma visão mais clara dos agrupamentos das variáveis nos fatores. Basicamente os carregamentos serão multiplicados por uma matriz ortogonal T sem prejudicar a reprodução da matriz de covariância em $\Sigma = LL' + \Psi$.

As communalidades h_i^2 , $i = 1, 2, \dots, p$ ficam inalteradas com a transformação $L^* = LT$. Denotando a i -ésima linha de L por λ_i' , a communalidade é dada por $h_i^2 = \lambda_i' \lambda_i$. A i -ésima linha de $L^* = LT$ é $\lambda_i'^* = \lambda_i' T$ e, transpondo-se ambos os membros, tem-se $\lambda_i'^* = (\lambda_i' T)' = T' \lambda_i$. Em termos de $\lambda_i'^*$, a communalidade é $h_i^{2*} = \lambda_i'^* \lambda_i'^* = \lambda_i' T T' \lambda_i = \lambda_i' I \lambda_i = \lambda_i' \lambda_i = h_i^2$. Neste trabalho optou-se pelo método de rotação de Kayser (Método Varimax Normal) [9].

2.6) Escores Fatoriais

Escores fatoriais são estimativas dos valores para os vetores fatoriais aleatórios não observáveis F_j , $j = 1, 2, \dots, n$. A estimação é complicada pelo fato de que as quantidades f_j e ε_j não são observadas como x_j . Como solução na estimação usou-se o Método dos Mínimos Quadrados Ponderados. Tomando-se $\hat{L}, \hat{\Psi}$ e $\hat{\mu} = \bar{x}$, o j -ésimo caso

$$\text{fica } \hat{f}_j = \left(\hat{L}' \hat{\Psi}^{-1} \hat{L} \right)^{-1} \hat{L}' \hat{\Psi}^{-1} (x_j - \bar{x}).$$

2.7) Coleta e Preparação dos Dados

Os métodos abordados foram aplicados a dados de Faxinais do Estado do Paraná, um tipo de Área Especialmente Protegida e próprio da Região Sul do Brasil. As informações foram obtidas por DOMINGUES [7], coletadas em 20 Faxinais pertencentes a 4 municípios do Estado do Paraná. E, as 80 variáveis que foram pesquisadas nos dois questionários do trabalho são: (1) proprietários; (2) arrendatários, parceiros ou agregados; (3) arrenda ou parceria; (4) renda média; (5) renda da agricultura; (6) renda de outras fontes; (7) áreas de lavoura; (8) áreas de floresta nativa; (9) áreas de floresta plantada; (10) áreas em descanso; (11) áreas no criadouro comunitário; (12) consumo de leite; (13) consumo de derivados de leite; (14) consumo de ovos; (15) consumo de carne de frango; (16) consumo de carne de gado; (17) consumo de carne de porco; (18) consumo de batata/mandioca; (19) consumo de farinha de trigo; (20) consumo de milho e derivados; (21) consumo de hortaliças; (22) consumo de legumes; (23) consumo de frutas; (24) consumo de doces; (25) consumo de óleo; (26) produção de erva-mate; (27) produtor de erva mate; (28) produtor de bracatinga; (29) consumo de lenha; (30) produtor de eucalipto para lenha; (31) plantam araucária; (32) mudas de araucária plantadas; (33) plantam erva mate; (34) mudas de erva mate plantadas; (35) plantam exóticas; (36) mudas de exóticas plantadas; (37) conservação de solos; (38) calagem de solo; (39) sementes selecionadas; (40) crédito agrícola; (41) assistência técnica; (42) renda de milho; (43) renda de feijão; (44) renda do fumo; (45) renda da erva mate; (46) número de bovinos; (47) número de animais de trabalho; (48) número de suínos; (49) número de galinhas; (50) filiado à associação; (51) participação em atividades comunitárias; (52) dias gastos em atividades comunitárias; (53) distância mínima da escola; (54) distância mínima do posto de saúde; (55) distância mínima do asfalto; (56) abastecimento de água com bomba elétrica; (57) rede pública comunitária de água; (58) luz de rede pública; (59) banheiro interno; (60) esgoto com fossa séptica; (61) sindicato para assistência médica; (62) rede pública para assistência médica; (63) tamanho da casa; (64) sempre moraram na comunidade; (65) uso múltiplo; (66) desenvolver e proteger; (67) futuro dos filhos no campo; (68) nunca pensaram em deixar de ser agricultor; (69) nunca pensaram em acabar com o Faxinal; (70) situação piora se o faxinal acabar; (71) criação solta como condição para faxinal; (72) produtores alfabetizados; (73) melhoramento genético da criação; (74) possui criadouro comunitário; (75) conhecimento do ICMS Ecológico; (76) possui associação; (77) possui escola; (78) possui posto de saúde; (79) densidade; (80) possui equipamentos comunitários.

Os dados que foram coletados por meio de questionários em entrevistas com 316 famílias amostradas em um universo de 1947 famílias, referem-se ao ano agrícola de agosto de 1997 a julho de 1998. A partir dos dados coletados, foi criada uma matriz de dados de ordem 20x80, sendo 20 o número de Faxinais considerados, e 80 o número de variáveis. Dessa matriz, calculou-se as correlações existentes entre cada par de variáveis, criando-se, a matriz de correlações. Em seguida, determinou-se os pares de autovalores e autovetores dessa matriz de correlação. Foram considerados 17 fatores que representam um grau de explicação de 97,7649%. A matriz de carregamentos fatoriais foi formada e os fatores foram, então, sem alteração do grau de explicação, rotacionados pelo Método Varimax Normal, sendo obtida, assim, a matriz de carregamentos fatoriais rotacionados. Em seguida, determinou-se as comunalidades e as variâncias específicas de cada variável. Os escores fatoriais, f_j , $j = 1, 2, \dots, n=20$, foram calculados pelo Método de Mínimos Quadrados Ponderados. A seguir calculou-se os escores classificatórios brutos E_j para os 20 faxinais ponderando-se os escores fatoriais originais pela importância de cada fator, representada pelo autovalor. Finalmente a escala de escores classificatórios E_j foi alterada para atender às exigências do IAP, que determina, pela Portaria n.º 263/98/IAP/GP de 1998, que a avaliação deve ter um mínimo de “0” e um máximo de “2”. Após esta padronização, foram encontrados os escores classificatórios finais.

A compreensão de todos os passos, desde os dados brutos, até os escores finais pode ser visualizada no Fluxograma (Figura 1).

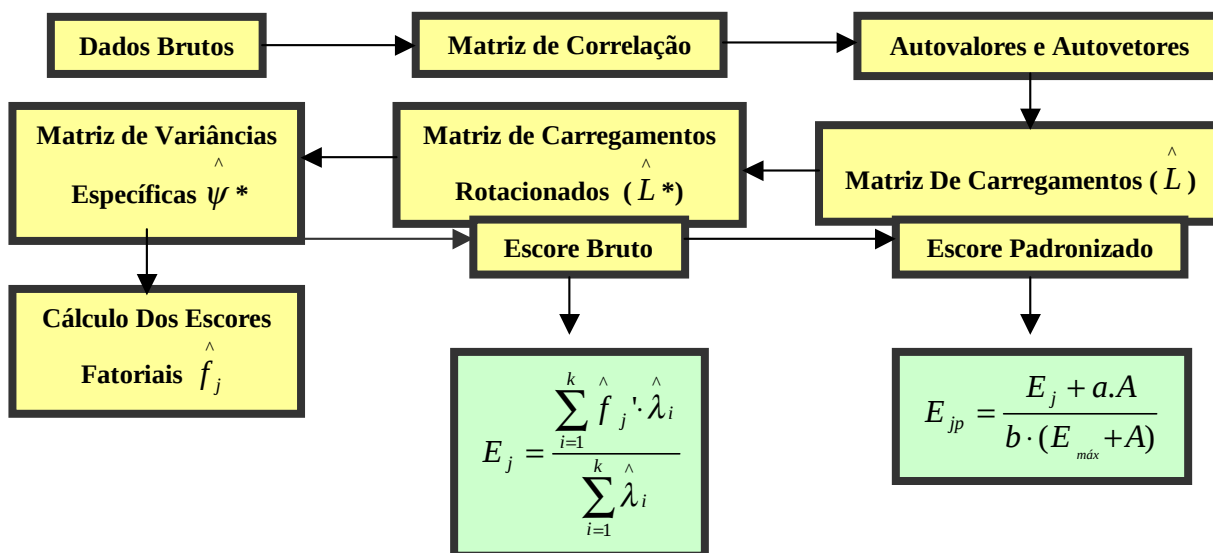


Figura 1 - Fluxograma do Cálculo do Escore Padronizado

II. RESULTADOS E CONCLUSÕES

Com a técnica utilizada, de acordo com o fluxograma anteriormente descrito, obteve-se a seguinte classificação:

Faxinal	Escore Bruto	Escore Padronizado	Faxinal	Escore Bruto	Escore Padronizado
1º São Pedro	406,0385	1,8	11º Tijuco Preto	-11,4835	0,8526
2º Ivaí-Anta Gorda	255,5732	1,4586	12º Paraná-Anta Gorda	-19,0331	0,8355
3º Ponte Nova	163,4198	1,2495	13º Guanabara	-71,2905	0,7169
4º Linha Brasília	150,4757	1,2201	14º Salto	-101,915	0,6474
5º Patos Velhos	116,3835	1,1427	15º Taboãozinho	-105,585	0,6391
6º Papanduva de Baixo	92,0459	1,0875	16º Dos Mellos	-113,344	0,6215
7º Queimadas	71,1442	1,0401	17º Marmeleiro de	-123,331	0,5988

			Baixo		
8° Cachoeira do Palmital	62,966	1,0215	18° Dos Krieger	-202,28	0,4197
9° Rio dos Couros	2,5779	0,8845	19° Marmeleiro de Cima	-264,043	0,2795
10° Rio do Meio	-9,2271	0,8577	20° Rio Bonito	-299,092	0,2

Essa classificação correspondeu a expectativa dos especialistas do IAP, ou seja, o Faxinal S. Pedro sempre foi considerado como o de melhor nível de conservação e o Faxinal do Rio Bonito é junto com o de Marmeleiro Grande aquele que apresenta maiores dificuldades na qualidade de vida dos seus moradores. O programa desenvolvido possibilitará, mediante atualização dos dados, classificações periódicas confiáveis baseadas em critério científico e não político.

IV. REFERÊNCIAS BIBLIOGRÁFICAS

- 1 ANDERSON, T. W. **An introduction to multivariate statistical analysis**, New York: John Wiley & Sons INC., 1958. 374 p.
- 2 BARTLETT, M. S. A note on multiplying factors for various chi squared approximations. **Journal of the Royal Statistical Society (B)**, 16, 1954. 296-298 p.
- 3 CHANG, M. Y. **Faxinais no Paraná**. Londrina: IAPAR, 1988. 20 p. (Informe de Pesquisa, n. 80).
- 4 _____. **Sistema Faxinal. Uma forma de organização camponesa em desagregação no Centro-Sul do Paraná**, Londrina: IAPAR, 1988, 123 p. (Boletim Técnico, n. 22).
- 5 CHAVES NETO. **Análise multivariada aplicada à pesquisa**. Curitiba: UFPR, 1998, 74 p. (notas de aula).
- 6 COOLEY, William W.; LOHNES, Raul R. **Multivariate data analysis**. New York: John Wiley & Sons INC. 1971. 364 p.
- 7 DOMINGUES, Zilna H. **Hierarquização de faxinais inscritos no cadastro estadual de unidades de conservação e uso especial, visando o ICMS ecológico**. Curitiba, 1999, Dissertação (Mestrado) – Setor de Ciências Agrárias, Curso de Engenharia Florestal, Universidade Federal do Paraná, 143 p.
- 8 JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. New Jersey: Prentice Hall, Englewood cliffs, 1998, 816 p.
- 9 KAYSER, Henry F. **The varimax criterion for analytic rotation in factor analysis**. Psychometrika, v. 23, n. 3, University of Illinois, 1958.
- 10 LAWLEY, D. N.; MAXWELL, A. E. **Factor analysis as a statistical method**. New York: American Elsevier Publishing Company, 1971, 153p.
- 11 LOUREIRO, Wilson. **ICMS Ecológico em perguntas e respostas**. Governo do estado do Paraná, 1998, 34p.