

## Análise da Distribuição de Bens Imobiliários no Brasil

Maria Cecília Bonato Brandalize, *PUCPR Curitiba PR*

Jürgen Philips, *UFSC Florianópolis SC*

1. Introdução
2. Análises Realizadas
  - 2.1 Análise de Cluster - Variáveis
  - 2.2 Análise Fatorial
3. Conclusão
4. Referências bibliográficas

### ▶ Resumo

## 1. Introdução

O Sistema Nacional de Cadastro Rural – SNCR, criado em 1972, tem por objetivo a promoção da integração e sistematização da coleta, pesquisa e tratamento dos dados e informações sobre o uso e a posse da terra. Isto significa, cadastrar e manter atualizados os dados sobre todos os imóveis rurais do país, entendendo como *Imóvel Rural* aquele de área contínua que seja ou possa ser destinado à exploração agrícola, pecuária, extrativa vegetal, florestal ou agro-industrial, independentemente de estar localizado na zona rural ou urbana de um município.

Assim, através da publicação de suas Estatísticas Cadastrais, o SNCR disponibiliza informações diversas sobre as condições vigentes no campo.

Tais informações, coletadas ao nível do Município, encontram-se totalizadas por Microrregião Demográfica, Unidades da Federação, Grandes Regiões e Brasil.

Numa primeira tentativa de análise dos dados utilizando a técnica de *Análise Fatorial, Método de Componentes Principais*, foram tomadas apenas sete variáveis (envolvendo somente valores de área em hectares) das mais de trinta e três disponibilizadas pelo Sistema (Tabela I a X do Volume I) e considerados apenas os valores totalizados para as Grandes Regiões (Norte, Nordeste, Sudeste, Sul e Centro-Oeste).

Em função do número de casos/indivíduos considerados (cinco Grandes Regiões) ter sido menor que o mínimo necessário (que é de seis) exigido para o processamento pelo software STATISTICA, a totalização por Grandes Regiões teve que ser abandonada e passou-se a considerar as totalizações por Unidade da Federação.

Alguns resultados relativos à distribuição e concentração de terras foram obtidos a partir da análise efetuada sobre as Unidades da Federação, porém, os mesmos se mostraram insatisfatórios, uma vez que não incluíam a frequência dos imóveis.

Numa segunda tentativa, outras variáveis consideradas importantes foram adicionadas à análise (totalizando quarenta, dentre as cento e sessenta e quatro disponibilizadas pelos Volumes I e II), envolvendo categorias variadas de área (em hectares) além das respectivas frequências dos imóveis.

Outros processamentos foram efetuados e, desta vez, em função do número de variáveis utilizadas, uma vez e meia superior ao número de casos/indivíduos, surgiram erros durante a análise (o determinante da matriz de correlação resultou igual a zero).

Para que a *Análise Fatorial pelo Método de Componentes Principais* pudesse ser realizada sem problemas, procedeu-se então a redução das variáveis significativas através de uma *Análise de Cluster*, também utilizando o software STATISTICA, o que exigiu a padronização das informações através da adoção de uma escala única de valores (conversão de todos os valores para porcentagem).

## 2. Análises Realizadas

Como explicitado anteriormente, as Estatísticas Cadastrais de 1998 apresentam, através de dez tabelas constantes no Volume I (de onde foram extraídos a maior parte dos dados) e de onze tabelas constantes no Volume II, uma quantidade bastante grande de informações relativas a 3.587.967 imóveis rurais cadastrados, totalizando 415.570.812,3 hectares que correspondem a 48,6% do território nacional.

A tabela a seguir apresenta as quarenta variáveis escolhidas e identificadas como sendo as mais importantes dentre as cento e sessenta e quatro que se encontravam disponíveis.

Variável	Descrição
AT	Área Total Cadastrada
ATANU	Área Total Aproveitável Não Utilizada
ATCEP	Área Total com Efetivo Pecuário
ATCEV	Área Total com Extração Vegetal
ATCUGA	Área Total com Utilização Granjeira ou Aquícola
ATE	Área Total Explorada
ATECC	Área Total Explorada com Cultura
ATGP	Área Total das Grandes Propriedades
ATINAP	Área Total Inaproveitável (Com/Sem Exploração Mineral)
ATINEX	Área Total Inexplorada
ATM	Área Total dos Minifúndios
ATMP	Área Total das Média Propriedades
ATPFB	Área Total de Imóveis de Pessoa Física Brasileira

<b>ATPJB</b>	<b>Área Total de Imóveis de Pessoa Jurídica Brasileira</b>
<b>ATPMA</b>	<b>Área Total de Preservação (Permanente + Reserva Legal)</b>
<b>ATPNC</b>	<b>Área Total das Propriedades Não Classificadas</b>
<b>ATPP</b>	<b>Área Total das Pequenas Propriedades</b>
<b>ATR</b>	<b>Área Total Registrada (Proprietário + Posseiro)</b>
<b>ATRCEN</b>	<b>Área Total Reflorestada com Essências Nativas</b>
<b>ATUCP</b>	<b>Área Total Utilizada com Pastagens</b>
<b>IANU</b>	<b>Total de Imóveis Aproveitáveis Não Utilizados</b>
<b>IC</b>	<b>Total de Imóveis Cadastrados</b>
<b>ICEP</b>	<b>Total de Imóveis com Efetivo Pecuário</b>
<b>ICEV</b>	<b>Total de Imóveis com Extração Vegetal</b>
<b>ICUGA</b>	<b>Total de Imóveis com Utilização Granjeira ou Aquícola</b>
<b>IE</b>	<b>Total de Imóveis Explorados</b>
<b>IECC</b>	<b>Total de Imóveis Explorados com Cultura</b>
<b>IGP</b>	<b>Total de Imóveis Classificados como Grande Propriedade</b>
<b>IINAP</b>	<b>Total de Imóveis Inaproveitáveis</b>
<b>IINEX</b>	<b>Total de Imóveis Inexplorados</b>
<b>IM</b>	<b>Total de Imóveis Classificados como Minifúndio</b>
<b>IMP</b>	<b>Total de Imóveis Classificados como Média Propriedade</b>
<b>IPFB</b>	<b>Total de Imóveis de Pessoa Física Brasileira</b>
<b>IPJB</b>	<b>Total de Imóveis de Pessoa Jurídica Brasileira</b>
<b>IPMA</b>	<b>Total de Imóveis com área de Preservação</b>
<b>IPNC</b>	<b>Total de Imóveis Não Classificados</b>

IPP	Total de Imóveis Classificados como Pequena Propriedade
IR	Total de Imóveis Registrados
IRCEN	Total de Imóveis Reflorestados com Essências Nativas
IUCP	Total de Imóveis Utilizados com Pastagens

As Unidades da Federação, por sua vez, constituem os Estados (RO, AC, AM, RR, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, SP, PR, SC, RS, MS, MT e GO) e o Distrito Federal (DF), totalizando vinte e sete os casos/indivíduos analisados.

Com base nestes dados é que foram realizadas as análises a seguir.

### 2.1 Análise de Cluster - Variáveis

A *Análise de Cluster* ou de Agrupamentos, segundo HAIR et. al. (1998), classifica *objetos* de tal forma que cada objeto é similar a outros dentro de um mesmo agrupamento (ou *cluster*), sendo esta similaridade dependente dos critérios preestabelecidos à análise.

Desta forma, os agrupamentos resultantes apresentam alta homogeneidade interna (entre os objetos de um mesmo *cluster*) e alta heterogeneidade externa (entre objetos de *clusters* distintos).

Tradicionalmente, durante uma análise de agrupamentos, tais *objetos* representam os casos ou indivíduos, sendo as variáveis utilizadas como representantes das características destes indivíduos.

Para este trabalho, no entanto, a técnica foi utilizada para agrupar variáveis, a fim de que se pudesse reduzi-las a um número considerado apropriado à *Análise Fatorial*.

De acordo com HAIR et. al. (1998), a *Análise Fatorial* exige que o número de observações (casos/indivíduos) seja, no mínimo, cinco vezes o número de variáveis a serem analisadas. Os mesmos autores recomendam que o número de observações da amostra utilizada não seja inferior a cinquenta, preferencialmente devendo ser superior a cem.

Ainda segundo HAIR et. al. (1998), o pesquisador deve tentar obter a maior razão de casos-por-variável possível, a fim de minimizar as chances de ajustar demais (*overfitting*) os dados. No entanto, amostras pequenas com menor razão de casos-por-variável podem ser analisadas, desde que a interpretação dos resultados seja feita com mais cuidado.

Como o número de casos/indivíduos da amostra considerada para o desenvolvimento deste trabalho é pequeno (apenas vinte e sete), o número de variáveis deve, necessariamente, sofrer uma redução. Esta redução, estipulada em um terço, pretende trabalhar treze das quarenta variáveis selecionadas elevando a razão de casos-por-variável para dois-para-um.

Assim, sobre os dados brutos tabelados no software STATISTICA (quarenta variáveis e vinte e sete indivíduos) procederam-se duas análises de *cluster* levando-se em consideração os seguintes critérios:

	1ª Análise	2ª Análise
Procedimento/Algoritmo	Hierárquico Aglomerativo	Hierárquico Aglomerativo
Medida de Similaridade	Distância Euclidiana	Distância Euclidiana ao Quadrado

Método

Ward's Method

Ward's Method

De acordo com HAIR et. al. (1998), a medida de similaridade (ou dissimilaridade) é feita utilizando-se distâncias. Assim, quanto mais próximos os objetos, maior a sua semelhança. Ainda segundo os mesmos autores, a medida de distância mais utilizada por esta técnica é a *Euclidiana*, simples ou elevada ao quadrado (distância simples sem a extração da raiz). Esta última é altamente recomendada quando a análise de agrupamentos é feita utilizando-se o *Método do Centróide* ou o *Método de Ward*.

O procedimento utilizado para o agrupamento das variáveis foi o *Hierárquico Aglomerativo*, onde os agrupamentos são hierarquizados através de uma estrutura em forma de árvore, que inicia com cada objeto fazendo parte de um *cluster*. Nas etapas seguintes, os *clusters* mais próximos são combinados dois a dois originando novos *clusters* agregados até que se tenha um único *cluster*, resultado das sucessivas agregações.

A representação gráfica desta estrutura recebe o nome de *Dendrograma*, da palavra inglesa *dendrogram*, com origem na palavra *dendrology*, ou dendrologia, relativamente ao tratado das árvores.

Dentre os mais populares métodos hierárquicos aglomerativos utilizados para o desenvolvimento de agrupamentos encontram-se (HAIR et. al., 1998):

- *Simple Linkage*: os *clusters* são construídos com base na distância mínima entre dois objetos (vizinho mais próximo).
- *Complete Linkage*: os *clusters* são construídos com base na distância máxima entre dois objetos (vizinho mais afastado).
- *Average Linkage*: os *clusters* são construídos com base na distância média tomada a partir de todos os objetos de um *cluster* em relação aos objetos de outro *cluster*.
- *Ward's Method*: distingue-se dos demais uma vez que utiliza a abordagem da análise de variância para avaliar a distância entre os *clusters*.
- *Centroid Method*: os *clusters* são construídos com base na distância entre os seus centróides.

O método escolhido para o agrupamento das variáveis, realizado através do software STATISTICA, foi o *Ward's Method*, pois este é considerado por vários autores como sendo bastante eficiente, embora tenha a tendência de criar *clusters* com o mesmo número reduzido de objetos.

BOUROCHE & SAPORTA (1980) indicam este método porque o critério de agrupamento utilizado satisfaz a condição de que a agregação de duas classes (*clusters*) só é realizada quando a perda de inércia é mínima.

Os resultados obtidos em função do processamento realizado na 1ª e 2ª análises, que diferem somente em relação à medida de similaridade empregada, foram:

Para a 1ª Análise – Utilizando a Distância Euclidiana:

- No dendrograma abaixo, a escala em Y (distância de ligação) varia de 0,00 a 2,14 para os 39 *clusters* formados. Percebe-se uma alta homogeneidade entre os 30 primeiros *clusters*, que encontram-se ligados por distâncias que variam entre 0,00 a 0,07.

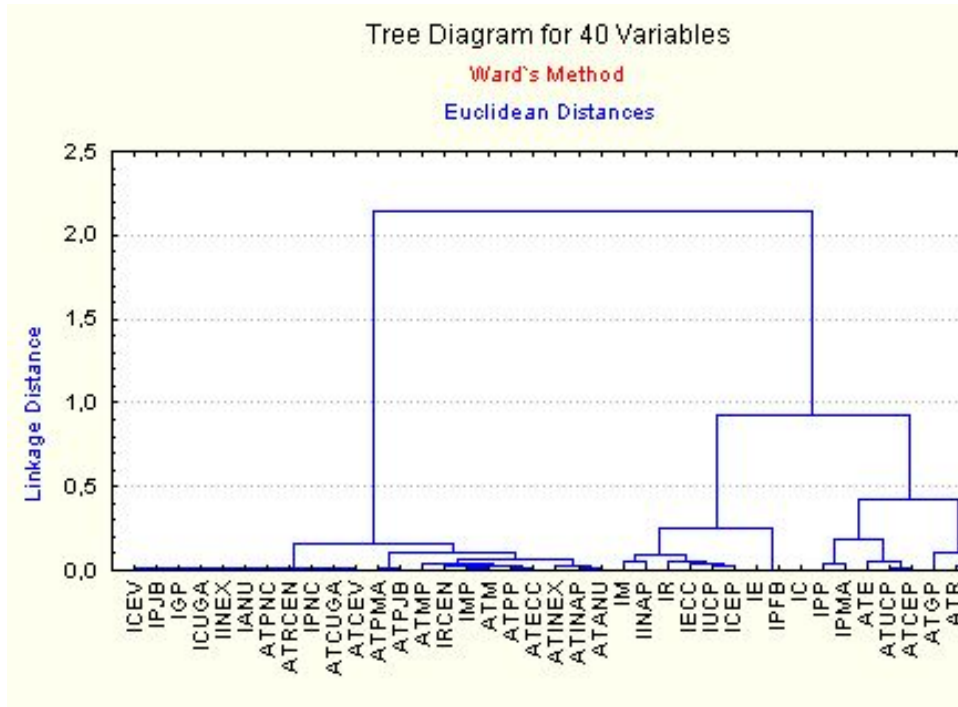


Figura 1

- Não foram detectados *outliers*, ou seja, objetos cujas observações não são representativas da população e normalmente identificados como ramos isolados da árvore e que acabam por encontrar os demais ramos somente quando próximos ao topo.
- Como o objetivo foi reduzir de quarenta para treze o número de variáveis, optou-se por cortar a árvore no nível correspondente ao valor 0,06 (distância de ligação), conforme apresentado no gráfico a seguir.

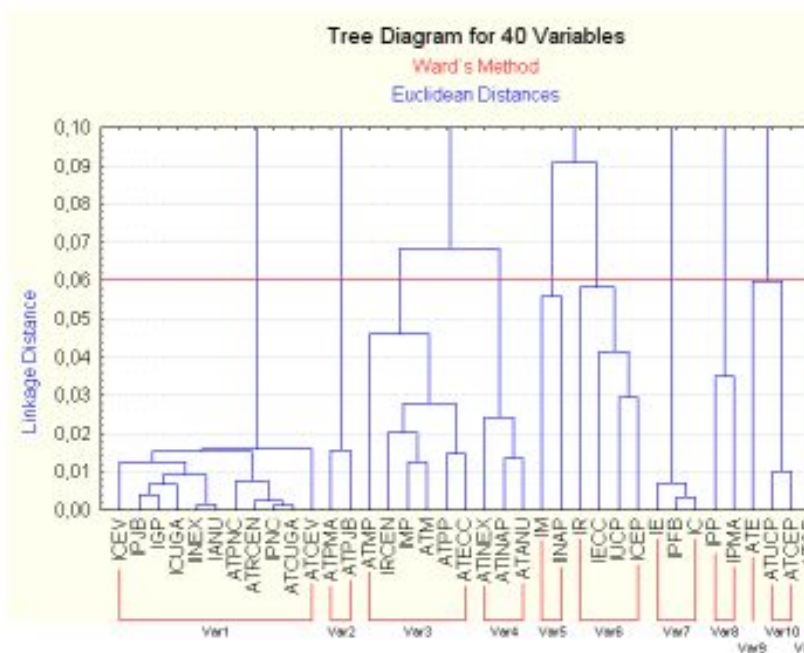


Figura 2

- De cada *cluster* observado abaixo do nível de corte foi extraída uma única variável representativa. São elas:

Var1	IINEX
Var2	ATPMA
Var3	IMP
Var4	ATANU
Var5	IM
Var6	IECC
Var7	IC
Var8	IPP
Var9	ATE
Var10	ATUCP
Var11	ATGP
Var12	ATR
Var13	AT

- O critério de escolha destas variáveis representativas, apesar de subjetivo, levou em consideração a proximidade entre elas (através da análise do *Amalgamation Schedule*) além do cuidado em selecionar variáveis com distintos significados. Por exemplo, o *cluster* número 2 compreende as variáveis ATPMA e ATPJB. Destas, escolheu-se a ATPMA para representá-lo (Var2). Por sua vez, o *cluster* número 8 compreende as variáveis IPP e IPMA. Destas, escolheu-se a IPP para representá-lo (Var8) pois a variável IPMA corresponde ao total de imóveis da variável ATPMA (relativa à área total), anteriormente escolhida como representante do *cluster* número 2.

Para a 2ª Análise – Utilizando a Distância Euclidiana ao Quadrado:

- No dendrograma abaixo, a escala em Y (distância de ligação) varia de 0,00 a 0,57 para os 39 *clusters* formados. Assim como para a análise anterior, percebe-se uma alta homogeneidade entre os 30 primeiros *clusters*, que encontram-se ligados por distâncias que variam entre 0,00 a 0,004.

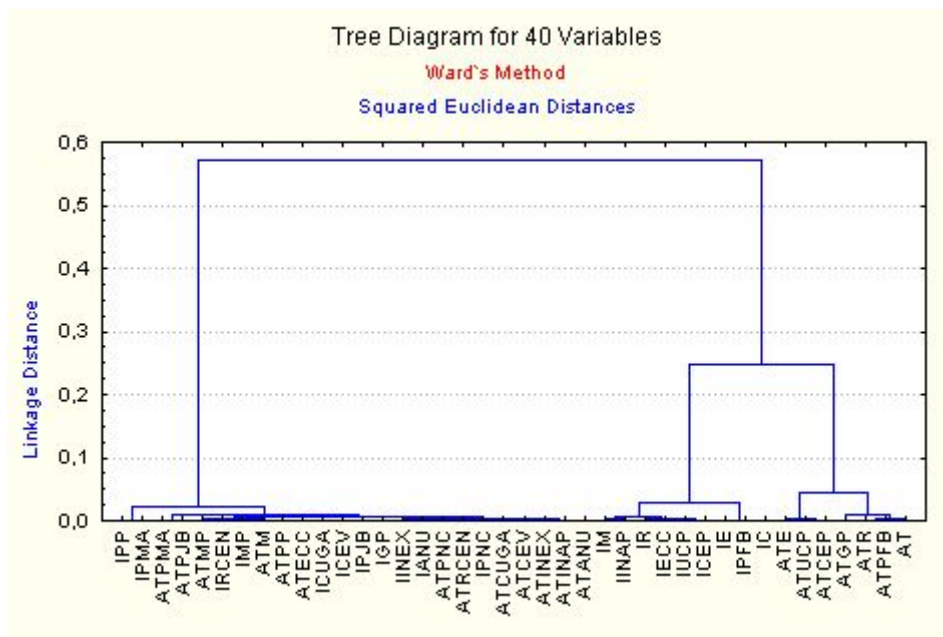


Figura 3

- Nesta análise também não foram detectados *outliers*.
- O corte da árvore, para a definição das treze variáveis, foi feito no nível correspondente ao valor 0,003 (distância de ligação), conforme apresentado no gráfico a seguir.

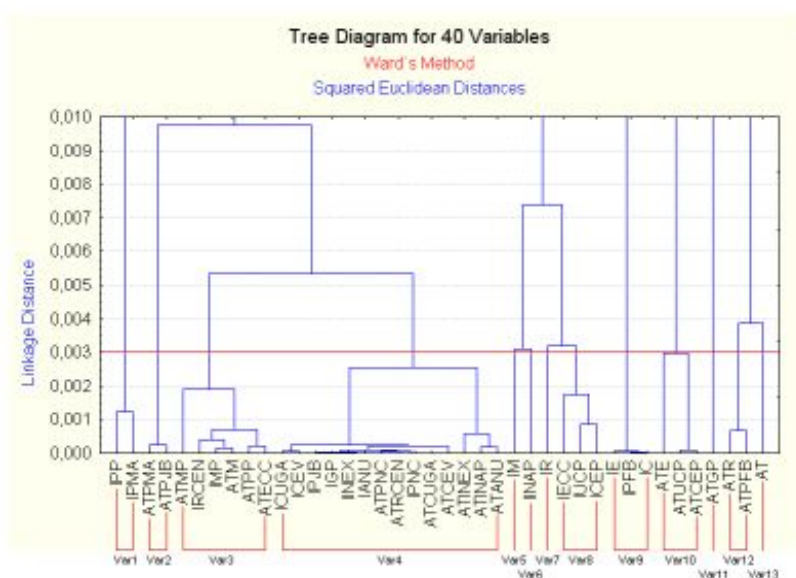


Figura 4

- Dos *clusters* localizados abaixo do nível de corte foram extraídas as seguintes variáveis representativas:

Var1	IPP
Var2	ATPMA



Var3	IMP
Var4	IINEX
Var5	IM
Var6	IINAP
Var7	IR
Var8	IECC
Var9	IC
Var10	ATUCP
Var11	ATGP
Var12	ATR
Var13	AT

- O critério utilizado para a escolha destas variáveis representativas foi o mesmo adotado para a análise anterior.

Percebe-se que o conjunto de variáveis resultante da primeira análise é bastante semelhante ao da segunda análise.

Para a *Análise Fatorial* considerou-se, no entanto, o conjunto resultante da primeira análise, uma vez que este apresenta maior equilíbrio das variáveis escolhidas relativamente às categorias de número de imóveis (seis variáveis) e de área (sete variáveis).

## 2.2 Análise Fatorial

A *Análise Fatorial*, segundo BOUROCHE & SAPORTA (1980), trata exclusivamente de variáveis numéricas e reduz o número destas variáveis não por uma simples seleção de algumas delas mas pela construção de novas variáveis sintéticas obtidas pela combinação linear das variáveis iniciais. Tal método permite representar, geometricamente, tanto os indivíduos quanto as novas variáveis obtidas.

De acordo com HAIR et. al. (1998), a *Análise Fatorial* é utilizada para determinar o relacionamento (correlação) entre um grande número de variáveis e se estas podem ser condensadas ou sumarizadas em um conjunto menor de componentes ou fatores.

O uso desta técnica pode ser tanto exploratório (definir uma estrutura para o conjunto de variáveis ou reduzir os dados) quanto confirmatório (testar hipóteses ou taxar o grau pelo qual os dados satisfazem uma estrutura predefinida).

A aplicação de tal técnica, neste trabalho, assumiu um caráter exploratório e objetivou estudar o comportamento da nuvem de pontos definida pelos imóveis e áreas (variáveis) em relação às unidades da federação (indivíduos).

Desta forma, sobre os dados reduzidos e tabelados no software STATISTICA (treze variáveis e vinte e sete indivíduos) procedeu-se a *Análise Fatorial*, através do *Método de Componentes Principais*, obedecendo os seguintes critérios:

	1ª Etapa	2ª Etapa
Número Máximo de Fatores	4	4
Autovalor Mínimo	0,5	0,5
Fator de Rotação	Sem Rotação	VARIMAX Normalizada

O *Método de Componentes Principais*, segundo HAIR et. al. (1998), é utilizado para condensar a maioria das informações originais por meio de um número mínimo de fatores.

O *Número Máximo de Fatores* expressa, portanto, o número máximo de dimensões (ou combinações) a serem extraídas do conjunto original de variáveis. Este valor, segundo os mesmos autores, é obtido pelo exame de um certo número de estruturas fatoriais derivadas de várias soluções, pois, só assim, o pesquisador pode comparar e contrastar as soluções encontradas para chegar à melhor representação dos dados.

O *Autovalor Mínimo* (conhecido como *Latent Root Criterion*) e também utilizado como critério para estabelecer o número de fatores a serem extraídos, baseia-se na idéia de que cada fator deve poder explicar pelo menos uma variável, à qual é atribuído o valor 1 na soma dos autovalores. Portanto, autovalores inferiores a um (1) não deveriam ser considerados significativos.

Para HAIR et. al. (1998), no entanto, este critério é válido somente quando o número de variáveis analisadas está entre 20 e 50, podendo ser alterado para as análises onde o número de variáveis analisadas é inferior a 20 ou superior a 50 e onde existe uma tendência do método em extrair um número, respectivamente, muito pequeno ou muito grande de fatores.

HAIR et. al. (1998) afirmam ainda que o pesquisador, numa primeira etapa, deve extrair os fatores sem rotação e, a partir da análise da matriz correspondente a estes fatores, explorar as possibilidades de redução dos dados para o conjunto de variáveis. Na etapa seguinte, os resultados devem ser rotacionados e o número final de fatores deve ser determinado para que os mesmos possam ser interpretados.

O *Fator de Rotação* é utilizado com o intuito de simplificar a estrutura fatorial e consequentemente a interpretação dos fatores. Nas soluções não rotacionadas a extração dos fatores é feita por ordem de importância, ou seja, o primeiro fator é extraído com base no cômputo da maior quantidade de variância enquanto os fatores subseqüentes são baseados no cômputo da variância residual (aquela que sobra do primeiro fator). Uma vez aplicada a rotação, têm-se como resultado uma redistribuição da variância entre os fatores.

Neste trabalho, os fatores foram rotacionados pelo método da VARIMAX Normalizada, que, dos métodos de rotação *ortogonal* (onde os eixos são mantidos à 90° indicando que os fatores continuam matematicamente independentes entre si), é o mais utilizado.

Assim, os resultados, a partir dos critérios adotados, foram:

Comuns à 1ª e 2ª Etapas:

- Número de Fatores Extraídos: 3
- Determinante da Matriz de Correlação: -21,307
- Autovalores:

Fator	Autovalor	% da Variância Total	Autovalor Acumulado	% Acumulada
1	7,372621	56,71247	7,372621	56,71247

2	3,969442	30,53417	11,34206	87,24664
3	0,909055	6,99273	12,25112	94,23937

Segundo HAIR et. al. (1998), quando a *Análise de Componentes Principais* é aplicada às Ciências Naturais, o número de fatores extraídos deve poder explicar uma *porcentagem de variância acumulada* igual ou superior a 95%, sendo que o último fator extraído deve poder explicar uma *porcentagem de variância total* superior a 5%.

Pelo quadro acima, os valores respectivos resultantes foram 94,24% e 6,99%. Isto implica que o número de fatores extraído é significativo.

- Este resultado pode ser confirmado através do gráfico apresentado a seguir, relativo ao critério do Autovalor Mínimo e ao SCREE TEST. Percebe-se que a partir do quinto fator a curva tende a horizontalizar-se (valores próximos de zero), indicando que os fatores correspondentes não são aceitáveis. No entanto, se fossemos considerar somente o critério do SCREE TEST, que define o número de fatores pelo primeiro ponto onde a curva inicia a sua horizontalização, então, o quarto fator também deveria ter sido incluído na análise (HAIR et. al., 1998).

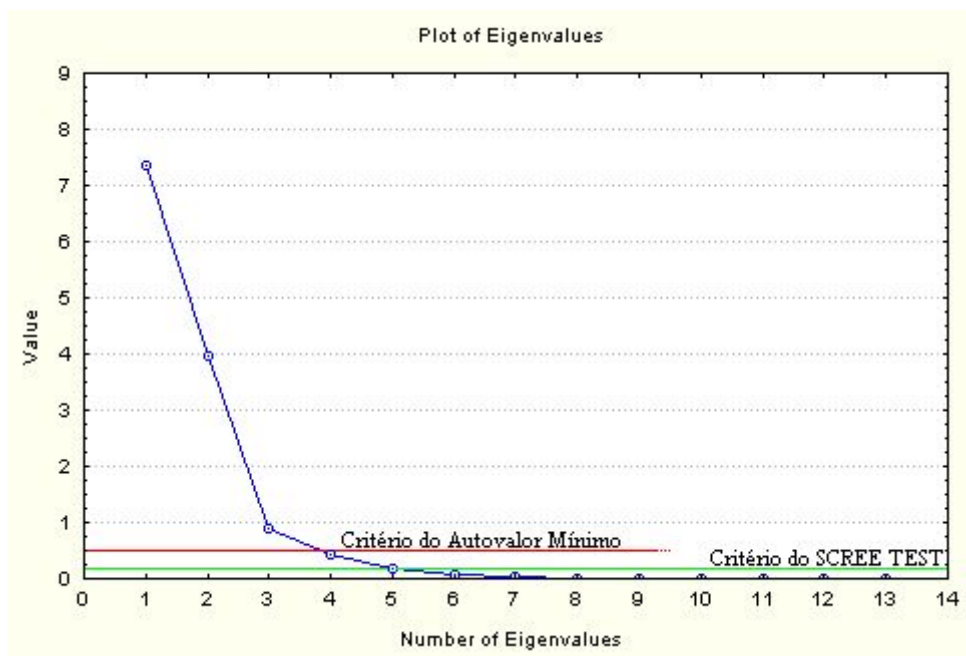


Figura 5

Relativos à 1ª Etapa: Sem Rotação

- Cargas Fatoriais:

Variável	Fator 1	Fator 2	Fator 3
IINEX	0,605074	0,057960	-0,657147
ATPMA	0,654972	0,698353	-0,086028
IMP	0,830980	-0,430777	0,144703
ATANU	0,393699	0,776455	-0,381574

IM	0,657879	-0,691478	-0,121491
IECC	0,673798	-0,708833	-0,101295
IC	0,722669	-0,672087	-0,096494
IPP	0,706719	-0,653637	-0,116300
ATE	0,947259	0,053641	0,285611
ATUCP	0,879962	0,137797	0,401349
ATGP	0,746406	0,653258	0,038320
ATR	0,891331	0,435534	0,088027
AT	0,894570	0,438299	0,059546

Na tabela acima, os valores representam as contribuições de cada variável para a formação dos fatores, ou seja, a correlação entre a variável original e o seu fator. Existem alguns critérios que determinam quais destes valores podem ser considerados significativos na formação de um determinado fator. Neste caso, o software STATISTICA automaticamente assinalou, em vermelho, os valores superiores a 0,7.

De acordo com HAIR et. al. (1998), cada variável deve ser assinalada ao menos uma vez, ou seja, cada variável deve estar melhor explicada por pelo menos um fator. As variáveis IINEX, ATPMA e IM, no entanto, não foram assinaladas pelo software (suas contribuições são inferiores a 0,7), estando suas cargas mais altas assinaladas em azul.

Quando isto ocorre, HAIR et. al. (1998) recomendam verificar a matriz de *comunalidades*. Nesta, os valores inferiores a 0,5 identificam as variáveis que não foram suficientemente explicadas pelos fatores e que, conseqüentemente, devem ser ignoradas ou mesmo eliminadas da análise.

O exame da matriz de *comunalidades*, no entanto, revelou valores superiores a 0,5 em pelo menos um dos fatores, para todas as variáveis. Desta forma, nesta etapa, nenhuma delas foi ignorada ou eliminada.

Relativos à 2ª Etapa: Com Rotação

- Cargas Fatoriais:

Variável	Fator 1	Fator 2	Fator 3
IINEX	0,276110	0,385629	0,759190
ATPMA	0,869963	-0,115811	0,392223
IMP	0,413316	0,848668	-0,077249
ATANU	0,626912	-0,320903	0,638343
IM	0,046929	0,958222	0,072906

IECC	0,055057	<b>0,980258</b>	0,052637
IC	0,114541	<b>0,982651</b>	0,067388
IPP	0,107155	<b>0,959723</b>	0,087560
ATE	<b>0,828512</b>	0,539696	-0,063620
ATUCP	<b>0,864424</b>	0,424070	-0,165341
ATGP	<b>0,951270</b>	-0,031841	0,281795
ATR	<b>0,947627</b>	0,224195	0,208924
AT	<b>0,942604</b>	0,226128	0,237215

Os novos valores das contribuições, calculados a partir da rotação pela VARIMAX Normalizada, permitem verificar que as cargas superiores a 0,7 encontram-se melhor distribuídas entre os fatores, relativamente aos resultados obtidos anteriormente.

No entanto, a solução ainda apresenta uma variável (ATANU) com carga máxima inferior a 0,7 (assinalada em azul). Os valores correspondentes à matriz de comunalidades, para esta variável, são: 0,4 em relação ao Fator 1; 0,5 em relação ao Fator 2; e 0,9 em relação ao Fator 3. Isto significa que a mesma se encontra melhor explicada pelo Fator 3, não devendo ser ignorada ou eliminada.

- Escores Fatoriais:

Caso	Fator 1	Fator 2	Fator 3
RO	-0,680814	-0,458521	0,97870
AC	-0,428142	-0,807007	-0,16871
AM	0,113307	-0,830751	0,30845
RR	-0,628592	-0,764878	0,91406
PA	1,255841	-0,911928	1,39704
AP	-0,542810	-0,769795	-0,58958
TO	0,601231	-0,559440	-1,16582
MA	-0,032053	-0,397102	0,12476
PI	-0,661565	-0,062205	1,72321
CE	-0,432526	-0,014655	-0,37192

RN	-0,553905	-0,529686	-0,46957
PB	-0,658043	-0,203966	-0,19133
PE	-0,644428	-0,074054	-0,18347
AL	-0,609880	-0,602943	-0,54576
SE	-0,633071	-0,531442	-0,50270
BA	0,414873	1,400930	0,28680
MG	1,287061	2,526610	-1,29645
ES	-0,538978	-0,279039	-0,50208
RJ	-0,603256	-0,411856	-0,34523
SP	-0,170303	1,408928	1,07791
PR	-0,519599	1,801738	0,70374
SC	-0,813910	0,755085	0,52620
RS	-0,300277	2,276854	0,27184
MS	1,828200	-0,518820	-1,99624
MT	3,349705	-0,776705	2,21324
GO	1,270419	0,065788	-1,18291
DF	-0,668484	-0,731137	-0,38421

Os valores de escores fatoriais apresentados na tabela acima representam o grau pelo qual cada indivíduo encontra-se relacionado ao grupo de variáveis explicado por um determinado fator. Isto significa que os escores fatoriais são computados com base nas cargas fatoriais de todas as variáveis que compõem um fator.

A análise destes valores, portanto, só tem sentido, se feita em conjunto à análise dos valores das cargas fatoriais.

Para melhor visualizar a relação existente entre as tabelas (cargas e escores) geram-se alguns gráficos, normalmente em duas dimensões, para cada par de fatores separadamente. Em seguida, os gráficos resultantes são combinados e interpretados.

- Gráficos Gerados:

Os gráficos bidimensionais gerados para interpretação dos resultados da análise fatorial foram:

- Cargas Fatoriais: Fator 1 x Fator 2; Fator 1 x Fator 3; e Fator 3 x Fator 2

- Escores Fatoriais: Fator 1 x Fator 2; Fator 1 x Fator 3; e Fator 3 x Fator 2

Interpretação Isolada dos Gráficos:

- *Cargas Fatoriais*: Fator 1 x Fator 2

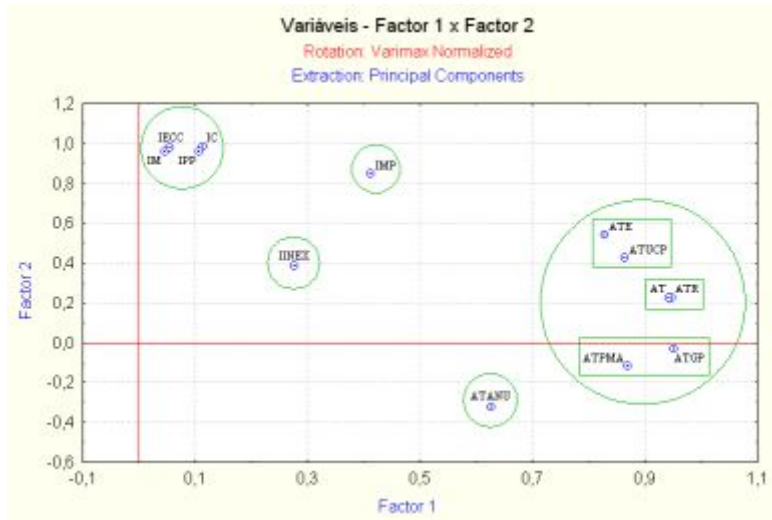


Figura 6

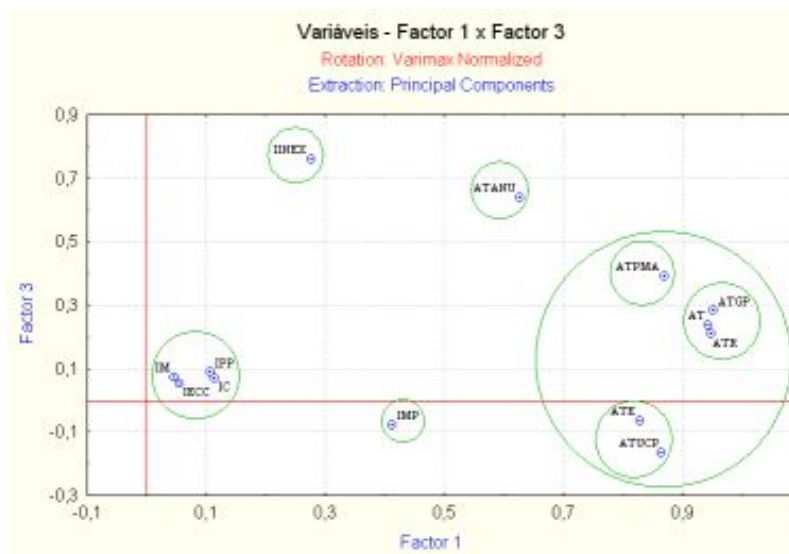
Este gráfico revela a forte oposição existente entre as variáveis de dois grupos bastante distintos. São eles:

**O grupo dos imóveis:** que concentra as pequenas propriedades (IPP) e os minifúndios (IM), indicando que estas propriedades são preferencialmente exploradas com culturas (IECC) além de representarem uma grande parcela dos imóveis cadastrados (IC).

**O grupo das áreas:** que concentra as grandes propriedades (AGP), indicando que sua principal utilização é a pastagem (ATUCP), pois provavelmente dedicam-se à exploração pecuária, além do que, são nestas propriedades que se encontram as maiores áreas destinadas à preservação do meio ambiente (ATPMA). As grandes propriedades representam ainda uma grande parcela da área total cadastrada (AT), das áreas exploradas (ATE) e das áreas registradas (ATR), ou seja, aquelas que efetivamente possuem título de propriedade.

É possível perceber ainda que as áreas aproveitáveis não utilizadas (ATANU) têm maior relação (estão mais próximas) com o grupo das áreas relativo às grandes propriedades (AGP), estando mais afastadas do grupo dos imóveis relativo às pequenas propriedades (IPP) e minifúndios (IM) e também dos imóveis classificados como médias propriedades (IMP).

- *Cargas Fatoriais*: Fator 1 x Fator 3



**Figura 7**

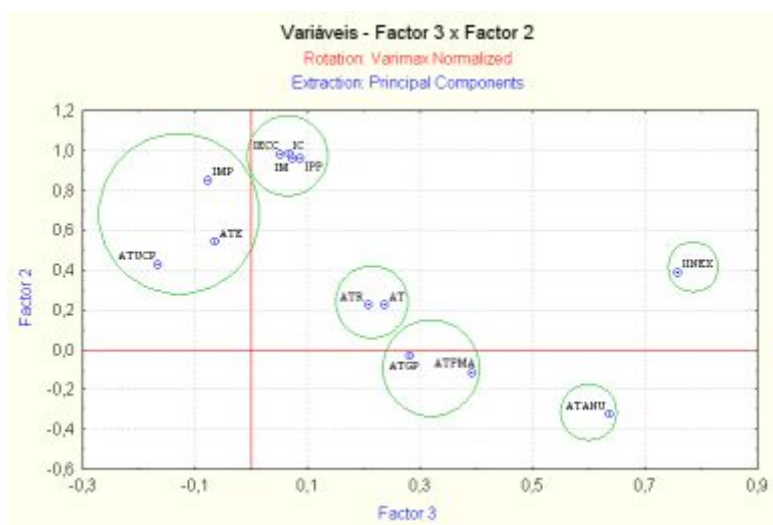
Mais uma vez, este gráfico revela a forte oposição entre o grupo dos imóveis (IPP, IM, IECC e IC) e o grupo das áreas (AT, ATGP, ATE).

Percebe-se ainda uma leve oposição entre as áreas destinadas à preservação do meio ambiente (ATPMA) e o grupo formado pelas áreas utilizadas com pastagens (ATUCP) e as efetivamente exploradas (ATE). Este tipo de oposição é óbvia, uma vez que as áreas destinadas à preservação do meio ambiente não podem sofrer qualquer tipo de exploração.

Confirma-se, em relação ao gráfico anterior, que as médias propriedades (IMP) se opõem às áreas aproveitáveis não utilizadas (ATANU).

A variável relativa aos imóveis inexplorados (IINEX) aparece isolada das demais.

- **Cargas Fatoriais:** Fator 3 x Fator 2



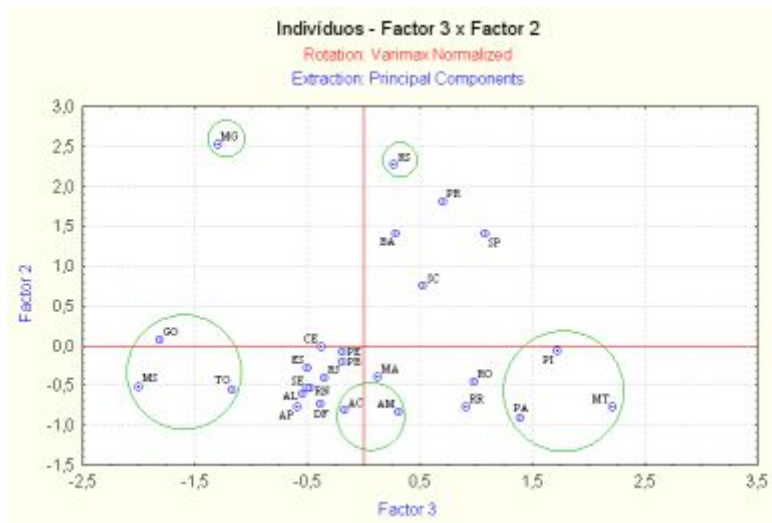
**Figura 8**

Este gráfico permite visualizar a relação existente entre o grupo dos imóveis (IPP, IM, IC e ICC), além daqueles classificados como média propriedade (IMP), e os imóveis inexplorados (IINEX), que é de forte oposição. Isto indica que tanto as pequenas como médias propriedades são altamente exploradas, o que se verifica pela sua proximidade com a variável ATE, além das variáveis ATUCP (pastagem) e IECC (cultura), relativamente ao tipo de utilização mais freqüente destes imóveis.





- **Escores Fatoriais: Fator 3 x Fator 2**



**Figura 11**

Este gráfico revela uma forte oposição entre o grupo formado pelos estados do Mato Grosso (MT), Piauí (PI) e Pará (PA) e o grupo dos estados formado por Mato Grosso do Sul (MS), Goiás (GO) e Tocantins (TO), todos em relação ao terceiro fator.

O gráfico revela ainda uma oposição entre o estado do Rio Grande do Sul (RS) e o grupo formado pelos estados do Acre (AC) e Amazonas (AM), relativamente ao segundo fator.

O estado de Minas Gerais (MG), mais uma vez, encontra-se isolado dos demais, estando melhor representado pelo segundo fator.

- **Interpretação Combinada dos Gráficos:**

Como os gráficos relativos às cargas fatoriais não possuem a mesma escala de valores dos gráficos relativos aos escores fatoriais, a interpretação conjunta destes torna-se mais complexa. Assim, para evitar erros grosseiros, procedeu-se a interpretação dos mesmos e os resultados foram checados junto às tabelas das Estatísticas Cadastrais de 1998, muitas vezes utilizada como referência para as descrições que se seguem.

É importante salientar que as comparações, e conseqüentemente as interpretações, foram feitas com base na posição que cada variável e indivíduo ocupam relativamente aos eixos tomados para projeção.

Assim, as interpretações descritas a seguir têm como base a análise da proximidade entre as variáveis e os indivíduos projetados, relativamente a cada um dos eixos do gráfico.

- **Cargas Fatoriais e Escores Fatoriais: Fator 1 x Fator 2**

**Em relação ao Fator 1:**

Nos estados do Mato Grosso (MT) e Pará (PA) concentram-se as áreas aproveitáveis mas não utilizadas (ATANU).

No estado do Mato Grosso (MT), onde a quantidade de área cadastrada (AT) é a maior do Brasil, encontra-se a maior quantidade de imóveis classificados como grande propriedade (ATGP), a maior quantidade de área registrada (ATR) e de área destinada à preservação do meio ambiente (ATPMA).

Nos estados do Rio Grande do Sul (RS) e Paraná (PR) concentram-se os imóveis classificados como pequena propriedade (IPP) e minifúndio (IM). Estes estados também são responsáveis por uma grande parcela dos imóveis cadastrados (IC), sendo explorados, predominantemente, com culturas (IECC).

**Em relação ao Fator 2:**

O estado de Minas Gerais (MG), que aparece isolado dos demais, é atípico, pois representa o estado com maior número de imóveis cadastrados (IC), perdendo em quantidade de área total

cadastrada (AT) somente para o estado do Mato Grosso (MT). Isto significa que, além de concentrar a maioria dos imóveis classificados como média propriedade (IMP) também concentra uma grande parcela dos imóveis classificados como pequena propriedade (IPP) e como minifúndio (IM). Este estado concentra, ainda, a maior quantidade de imóveis explorados com culturas (IECC), além da maior área utilizada com pastagens (ATUCP), sendo, portanto, o estado com maior área total explorada (ATE) do Brasil.

De acordo com os gráficos, uma grande parcela dos imóveis inexplorados (IINEX) está concentrada nos estados da Bahia (BA) e de São Paulo (SP).

- *Cargas Fatoriais e Escores Fatoriais: Fator 1 x Fator 3*

#### Em relação ao Fator 1:

Estes gráficos confirmam as interpretações realizadas para o estado do Mato Grosso (MT), relativamente ao item anterior.

Ainda em relação ao estado do Mato Grosso (MT), os gráficos revelam que este possui uma forte relação com a quantidade de área explorada (ATE) e utilizada com pastagem (ATUCP), representando, segundo as tabelas das Estatísticas Cadastrais de 1998, o segundo maior estado nestas categorias.

Os gráficos permitem identificar ainda que os estados do Mato Grosso do Sul (MS) e de Goiás (GO), depois dos estados de Minas Gerais (MG) e Mato Grosso (MT), representam os estados que mais concentram áreas exploradas (ATE) utilizadas com pastagens (ATUCP).

#### Em relação ao Fator 3:

A quantidade de imóveis inexplorados (IINEX), por sua vez, demonstra forte relação com os estados do Piauí (PI) e do Pará (PA). E isto se verifica pela consulta às tabelas das Estatísticas Cadastrais de 1998, onde este último aparece como o maior detentor de áreas inexploradas (incluindo as aproveitáveis mas não utilizadas, as inaproveitáveis e as de preservação do meio ambiente).

- *Cargas Fatoriais e Escores Fatoriais: Fator 3 x Fator 2*

#### Em relação ao Fator 3:

Os estados do Amazonas (AM), Roraima (RR) e Pará (PA), juntos, concentram as maiores propriedades (ATGP) do país. Para se ter uma idéia, a média de hectares por imóvel no Amazonas é de 19,6 mil, enquanto no Pará e em Roraima esta mesma média cai para 7,2 mil (valores calculados a partir das tabelas constantes das Estatísticas Cadastrais de 1998).

O estado do Pará (PA), depois do estado do Mato Grosso (MT), possui a maior quantidade de área destinada à preservação do meio ambiente (ATPMA).

O estado do Mato Grosso (MT) é o que mais concentra imóveis inexplorados (IINEX) no país.

#### Em relação ao Fator 2:

Os gráficos confirmam as interpretações realizadas para o estado de Minas Gerais (MG), relativamente ao primeiro item.

O estado de Santa Catarina (SC), cuja projeção no gráfico dos Escores Fatoriais corresponde à projeção das variáveis AT (área total cadastrada) e ATR (área total registrada) no gráfico das Cargas Fatoriais, não possui as maiores concentrações destas áreas, no entanto, possui a maior razão entre elas, ou seja, o total de áreas registradas relativamente ao total de áreas cadastradas, para este estado, chega a 95,6% (valor calculado a partir das tabelas constantes das Estatísticas Cadastrais de 1998).

### **3. Conclusão**

É evidente que a Análise Fatorial realizada sobre os dados facilitou bastante o processo de interpretação dos mesmos, uma vez que a quantidade de informações que se dispunha era bastante grande.

No entanto, com a redução drástica das variáveis, de cento e sessenta e quatro para quarenta e depois para apenas treze, a sensação é de que algumas informações importantes deixaram de ser contempladas pela análise.

Uma solução seria ampliar a amostra ao nível do município e, desta forma, ter-se-ia a razão ideal de casos-por-variável proposta por HAIR et. al. (1998) e que resultaria em trinta e três casos para cada variável, relativamente aos 5.506 municípios do país e às 164 variáveis disponíveis.

No entanto, com um número tão grande de casos e de variáveis, a interpretação dos resultados, de tão complexa, talvez se tornasse impossível.

Outro problema encontrado durante a análise está relacionado ao software utilizado, o STATISTICA, bastante limitado no que diz respeito à edição dos gráficos. Assim, os gráficos resultantes tiveram de ser exportados no formato *bitmap* para serem editados em outro software (PAINTSHOP PRO), demandando um tempo extra considerável.

Mesmo assim, os resultados obtidos através das análises efetuadas foram bastante interessantes, não restando dúvidas quanto à eficiência dos métodos utilizados.

A conclusão a que se chega, portanto, é bastante simples: os métodos de análise multivariada são capazes de produzir resultados surpreendentes, desde que o pesquisador conheça, suficientemente, os dados de que dispõe, os métodos apropriados à análise destes dados e o software utilizado para o processamento.

#### 4. Referências Bibliográficas

- BOUROCHE, J.-M.; SAPORTA, G.: *Análise de Dados*. Zahar Editores, 115p, Rio de Janeiro 1982  
HAIR, F.J.; et.al.: *Multivariate Data Analysis*, 5<sup>th</sup> Edition, 730p., Prentice Hall, New Jersey 1998  
INCRA, Instituto Nacional de Colonização e Reforma Agrária. *Estatísticas Cadastrais - Ano de 1998 - Volume I*, INCRA/DC, Brasília 1999  
INCRA, Instituto Nacional de Colonização e Reforma Agrária. *Estatísticas Cadastrais - Ano de 1998 - Volume II*, INCRA/DC, Brasília 1999

<b>Autores</b>	Maria Cecília Bonato Brandalize, Doutoranda na Universidade Federal de Santa Catarina - UFSC Departamento de Engenharia Civil Curso de Pós-Graduação em Cadastro Técnico Multifinalitário e Gestão Territorial 88040-900 Florianópolis SC - Brasil ✉ <a href="mailto:mcbonato@rla01.pucpr.br">mcbonato@rla01.pucpr.br</a>	Prof. Dr.-Ing. Jürgen Philips Universidade Federal de Santa Catarina - UFSC Departamento de Engenharia Civil Curso de Pós-Graduação em Cadastro Técnico Multifinalitário e Gestão Territorial 88040-900 Florianópolis SC - Brasil ✉ <a href="mailto:philips@ecv.ufsc.br">philips@ecv.ufsc.br</a>
----------------	--	--