

Identificação e Elaboração de Variáveis Estatísticas Através de Técnicas Multivariadas no Estudo da Produção de Grãos e Fibras, na Região Administrativa de Presidente Prudente-SP

Identification and Elaboration of Statistical Variables Through Multivariate Techniques in the Study of the Production Grain and Fiber, in Administrative Region of Presidente Prudente-SP

Eduardo Cardoso de Oliveira¹; Antonio Assiz de Carvalho Filho²; Ricardo Firetti³

Resumo

Este trabalho trata do problema para obtenção de variáveis estatísticas através de técnicas multivariadas aplicando aos dados da produção de grãos e fibras na Região Administrativa de Presidente Prudente-SP, e fundamentado no fato da região ser beneficiada por possuir um sistema logístico de distribuição trimodal (ferroviário hidroviário e rodoviário). O banco de dados utilizado foi extraído do Levantamento de Unidades de Produção Agropecuárias (LUPA) e refere-se às áreas disponibilizadas pelos municípios estudados para a produção de grãos e fibras e também de variáveis econômicas relacionadas ao agronegócio. As técnicas estatísticas aplicadas foram análise fatorial e análise de agrupamentos. Através da análise fatorial foram obtidas, das 13 variáveis originais, três descritores estatísticos (fatores). Os agrupamentos foram feitos a partir dos escores fatoriais, encontrando-se três grupos por descritor e também no conjunto de descritores. Ao final do trabalho, caracterizou-se a 10ª R.A. do Estado de São Paulo segundo as variáveis selecionadas todas do ano de 2008, dando embasamento para auxiliar na elaboração de estratégias empresariais e de formulação de políticas públicas, para o desenvolvimento da região em estudo.

Palavras-chave: Análise Fatorial, Seleção de variáveis, Programa R.

Abstract

This work deals with the study to obtain statistical variables through multivariate techniques applied to the data of grain and fiber production in the Administrative Region of Presidente Prudente-SP, and is based on the fact that the region be benefited by having a distribution logistics system three-ways (rail, road and waterway). The used database was extracted from the Survey of Agricultural Production Units (LUPA) and refers to the areas provided by the municipalities studied for the production of grains and fiber and also economic variables related to agribusiness. The applied statistical techniques were factor

¹ Aluno de mestrado do Departamento de Matemática e Ciências da Computação e Departamento de Estatística, Universidade Estadual Paulista “Júlio de Mesquita Filho”; edcarov@terra.com.br

² Professor titular da Faculdade de Ciências e Tecnologia da UNESP - Departamento de Estatística, Universidade Estadual Paulista “Júlio de Mesquita Filho”; assiz@fct.unesp.br

³ Pesquisador da Agência Paulista de Tecnologia do Agronegócios (APTA/SAA-SP); rfiretti@apta.sp.gov.br

analysis and cluster analysis. Through factorial analysis were obtained, from 13 original variables, three statistical descriptors (factors). The clusters were made from the scores factorials, finding three groups per descriptor and also by the conjoint of the descriptors. At the end of the work, was characterized the 10th RA State of São Paulo according to the variables selected all of the year 2008, giving basement to auxiliary in elaboration business strategies and formulation of the public policies for the development of the region under study.

Key words: Factor Analysis, Selection of variables, R Program.

Introdução

De acordo com Micheletto (2009), a área plantada de grãos e fibras, no Brasil no período de 1985 a 1989, foi de 38,5 milhões de hectares, em média, envolvendo as culturas arroz, soja, algodão, milho trigo e feijão. Sendo que na safra de 2008/09, plantou-se 46,2 milhões. A expansão da área plantada de grãos e fibras nos últimos 20 anos foi de apenas 20%, indicando um crescimento anual médio em torno de 0,6% ao ano.

A cultura de soja foi a única que apresentou desenvolvimento expressivo, com acréscimo de 122% no mesmo período sustentando o desempenho global da área plantada no país. A área plantada de grãos e fibras no país cresceu apenas 7,7 milhões de hectares. Com base nesses dados as outras atividades tiveram redução na área de plantio.

Segundo Micheletto (2009), se for observar o crescimento dos números registrados nas estatísticas oficiais envolvendo o plantio de área de grãos e fibras, com certeza a floresta amazônica não está sofrendo devastação. A produção agrícola brasileira está obtendo grande ganho com o aumento de produtividade devido à incorporação de tecnologia, cujo resultado econômico está sendo transferido principalmente para o consumidor e para o comércio externo. Uma parcela muito pequena está ficando com os produtores, tendo em vista o elevado grau de descapitalização e de endividamento em que se encontram. A “porteira”, no momento, está fechada para o agricultor obter lucro.

O crescimento das áreas cultivadas com algodão, feijão, soja e arroz, afetado principalmente pela menor influência do fenômeno La Niña sobre

essas culturas, levou a Companhia Nacional de Abastecimento (Conab) a projetar uma colheita de 153 milhões de toneladas de grãos na safra 2010/2011. Segundo a estatal, a má distribuição das chuvas foi menos prejudicial do que o esperado.

O valor estimado pela Conab no quinto levantamento para esta safra, consolida a possibilidade de um novo recorde de produção, com aumento de 2,6%, ou 3,8 milhões de toneladas em relação ao ciclo passado, de 149,2 milhões de toneladas. Na comparação com o levantamento anterior, anunciado há um mês, o crescimento foi de 2,4%, ou 3,6 milhões de toneladas.

O quinto levantamento da safra de grãos 2010/2011 foi realizado entre os dias 16 e 21 de janeiro por 58 técnicos da Conab. Eles colheram informações com representantes de cooperativas e sindicatos rurais, órgãos públicos e privados nas regiões Sul, Sudeste e Centro-Oeste, além de parte do Norte e Nordeste.

Considerando-se então os fatos de que a Região Administrativa (RA) de Presidente Prudente tem boa localização geográfica no que diz respeito à logística de distribuição e tem sua estrutura produtiva com forte perfil agroindustrial, a realização deste estudo, ao destacar como de importante relevância a análise da produção de grãos e fibras, se constituirá como suporte às ideias apresentadas em Barat et al. (2007).

A logística, em sentido amplo, e o transporte, como atividades-meio de apoio à produção de bens e serviços, são essenciais para o aproveitamento do potencial produtivo de uma região, país ou associação de países. Estão na linha de frente de questões ligadas ao desenvolvimento econômico e

social, à geração de empregos, à ocupação territorial, à energia e ao meio ambiente (BARAT et al., 2007, p. 41). Dessa forma, têm profundas implicações no processo de desenvolvimento de uma região.

Para a análise dos dados serão utilizadas técnicas estatísticas multivariadas, pois se deseja formar novas variáveis capazes de auxiliar na interpretação de como se comporta a produção de grãos e fibras na região em questão. Os dados que estão sendo trabalhados são as áreas em produção referentes ao ano de 2008, e foram coletados no armazém de dados oficial do Instituto de Economia Agrícola (IEA).

Com o apoio da Agência Paulista de Tecnologias dos Agronegócios (APTA/SAA-SP), Pólo da Alta Sorocabana sediado em Presidente Prudente, ao final desse trabalho espera-se que identificar qual a relevância dessa produção para a RA de Presidente Prudente, e assim fornecer subsídios ao setor público e setor privado para formulações de propostas que influenciem no desenvolvimento da região em questão.

Procedimentos metodológicos

Este trabalho teve como tema as áreas destinadas à cultura de grãos e fibras na 10ª Região Administrativa do Estado de São Paulo (Região Administrativa de Presidente Prudente), disponíveis em armazéns de dados da Fundação SEADE e do censo do Instituto de Economia Agrícola, o Levantamento de Unidade de Produção Agropecuária (LUPA) do estado de São Paulo, referentes ao ano de 2008. Utilizou-se para tanto uma combinação de métodos de pesquisa descritiva e exploratória multivariada.

A Região em Estudo

A Região Administrativa de Presidente Prudente situa-se no extremo oeste do Estado de São Paulo, é a quarta maior em extensão territorial, com 23.952 km², ocupando 9,6% do território paulista. Nela se localizam 53 municípios distribuídos em

três regiões de governo: Adamantina, Dracena e Presidente Prudente (SEADE, 2007).

No ano de 2001, a região era responsável por cerca de 5,3% da produção agropecuária do Estado de São Paulo. Além dessa importante contribuição, a região responde por cerca de 18% da produção total de carne bovina no Estado, garantindo assim, a posição de maior exportadora desse produto em âmbito nacional, além de ser considerada uma das maiores bacias leiteiras do país. Podemos adicionar ainda, a produção de cana-de-açúcar e de ovos, traçando assim, o perfil agropecuário da região.

Essa Região Administrativa é favorecida por uma boa condição de escoamento de produtos, pois ela possui um sistema de logística trimodal, ou seja, Fluvial, Rodoviário e Ferroviário, fato que cria condições de mobilidade e fluidez dos produtos produzidos na região.

No que diz respeito ao modal Fluvial, a região conta com o Porto de Presidente Epitácio, construído na década de 60, com o objetivo de interligar o Rio Paraná com o terminal ferroviário da antiga Companhia Paulista de Estradas de Ferro. Decorrente do alagamento de boa parte da área próxima ao Porto para a construção da Usina Hidrelétrica Sérgio Motta, este foi desativado em 1998. Através de um acordo com a Companhia de Energia Paulista (CESP), começou então um trabalho de reconstrução e suas atividades portuárias foram retomados em abril de 2008.

Além de estar na área de influência da Hidrovia Tietê-Paraná, que forma um corredor de escoamento de produtos para países do Mercosul, a Região Administrativa de Presidente Prudente tem como principais vias de acesso à Rodovia Raposo Tavares (SP-270), que faz a ligação com a capital paulista e o Estado de Mato Grosso do Sul; a Assis Chateaubriand (SP-425), que conecta ao Paraná e às regiões do oeste e norte do Estado de São Paulo; a Comandante João Ribeiro de Barros (SP-294), que faz a junção de vários municípios da RA com

o Rio Paraná e, no sentido oposto, com Bauru e a Rodovia Marechal Rondon (SEADE, 2007).

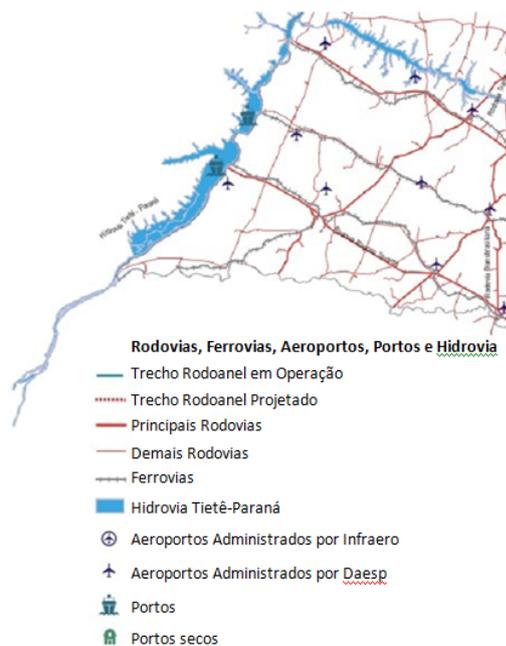
Os trilhos da Estrada de Ferro Sorocabana, hoje ALL - América Latina Logística chegaram a Presidente Prudente no dia 19 de janeiro de 1919, sendo inaugurada também a estação de Regente Feijó. Três anos depois a ferrovia atingiu a cidade de Presidente Epitácio, às margens do Rio Paraná.

Com a chegada das ferrovias houve uma grande expansão na economia, principalmente pelo fato da interligação entre a ferrovia da Alta Sorocabana com o porto de Presidente Epitácio, Paranaguá e o Porto de Santos, favorecendo o escoamento da produção agrícola.

No cenário atual, das ferrovias existentes na região, grande parte está desativada e sem qualquer possibilidade de serem reinstaladas, e as que ainda funcionam encontram-se abandonadas.

Na Figura 1 abaixo, segue o desenho das rodovias, hidrovias e ferrovias que cruzam a 10ª RA do Estado de São Paulo.

Figura 1 - Recorte da rede de transportes da 10ª RA do estado de São Paulo, 2005.



Fonte: Departamento Nacional de Infra-estrutura de Transporte, 2005.

Técnicas Estatísticas Multivariadas

Segundo Mingoti (2005), a estatística multivariada pode ser dividida em dois grupos: o primeiro diz respeito às técnicas exploratórias e de sintetização da estrutura dos dados e o segundo diz respeito aos métodos de estimação de parâmetros, testes de hipóteses, análise de variância, de covariância e de correlação multivariados.

Neste trabalho, serão estudadas apenas técnicas de caráter exploratório.

Análise Fatorial

Segundo Hair et al. (2005), a análise fatorial é uma técnica de interdependência, ou seja, na qual todas as variáveis são simultaneamente consideradas, cada uma relacionada com as outras, empregando ainda o conceito da variável estatística, a composição linear de variáveis.

Essa técnica busca condensar a informação contida em um número de variáveis originais em um conjunto menor de variáveis alternativas, chamadas variáveis latentes ou fatores, de forma que se perca o mínimo de informação possível. Tais fatores são formados a fim de maximizar o poder de explicação de todo conjunto de variáveis.

“A análise fatorial foi introduzida por Sperman (1904). Muitos dos exemplos iniciais aparecem na área de psicologia e ciências sociais, na tentativa de identificar os fatores relacionados com a inteligência humana e ligá-los, de algum modo, à etnia”. (MINGOTI, 2005, p. 100). Desde então a aplicação dessa técnica vem sendo praticada em diversas áreas como química, geologia, marketing, economia, agropecuária, pesquisa de mercado, entre outras.

De acordo com os objetivos do pesquisador, as técnicas fatoriais podem ser abordadas por meio de duas perspectivas:

- **Confirmatória:** Baseando-se em estudos teóricos ou pesquisas já realizadas, o pesquisador já tem a princípio alguma ideia sobre a real estrutura

dos dados, ou seja, ele tem um modelo fatorial hipotético pré-especificado ao qual deseja confirmar se realmente é aplicável.

- Exploratória: Nesse caso busca-se apenas encontrar uma estrutura subjacente ao conjunto de variáveis ou simplesmente a redução dos dados, sem nenhuma pré-suposição.

Ainda pode-se classificar a análise fatorial em tipo R, que é o mais utilizado e visa identificar as dimensões latentes através da análise de um conjunto de variáveis, ou tipo Q que agrupa populações em subpopulações.

De acordo com Hair et al. (2005), os desvios de normalidade, homoscedasticidade e linearidade devem ser levados em consideração para a aplicação da análise fatorial à medida que diminuem as correlações entre as variáveis em questão. A suposição de normalidade dos dados é exigida somente quando se deseja aplicar testes para a significância de fatores.

No que diz respeito às correlações da matriz de dados, recomenda-se segundo Hair et al. (2005) um número substancial de correlações maiores que 0,30; caso contrário a análise fatorial se torna inapropriada. Outra forma de analisar as correlações é através da matriz anti-imagem, que nada mais é que o valor negativo das correlações parciais, mas nesse caso os valores devem ser baixos.

Outra forma de garantir a adequação dessa técnica multivariada é examinando a matriz de correlação inteira através do teste Bartlett de esfericidade. Este teste fornece a probabilidade de haver correlações significantes pelo menos entre algumas das variáveis.

O teste de Alpha de Cronbach também é uma ferramenta de grande utilidade para avaliar as correlações. Segundo literatura estudada, ele fornece

uma medida de confiabilidade que varia de 0 a 1, sendo que valores de 0,6 a 0,7 são considerados o limite inferior de aceitabilidade. Em síntese, o teste proporciona medir a correlação existente entre as variáveis, podendo sofrer distorção em decorrência do número de variáveis envolvidas (não se aconselha o emprego de menos de 5 variáveis para realização do teste).

Outra medida que pode ser utilizada para quantificar o grau de intercorrelações entre as variáveis é a medida de adequação da amostra (MSA), que varia entre 0 e 1, sendo que quanto mais próximo de um melhor é a previsão sem erro de cada variável pelas outras variáveis. De acordo com Barroso e Artes (2003), a medida MSA pode ser interpretada com as orientações presentes na Tabela 1:

Tabela 1 - Valores de referência para a interpretação do MSA¹

Valores do MSA	Interpretação
$0,00 \leq MSA \leq 0,50$	Insuficiente
$0,50 \leq MSA \leq 0,60$	Regular
$0,60 \leq MSA \leq 0,70$	Bom
$0,70 \leq MSA \leq 0,80$	Ótimo
$0,80 \leq MSA \leq 1,00$	Excelente

Fonte: Barroso e Artes, 2003.

Existem vários métodos para a estimação das matrizes, tais como: Método dos componentes principais, Método dos fatores principais, Método iterativo dos fatores principais, Método da máxima verossimilhança e Método dos fatores canônicos.

O método dos componentes principais é usado quando se deseja encontrar o menor número possível de fatores necessários para explicar a maior parte da

¹ O MSA também pode ser obtido pelo teste de Kaiser-Meyer-Olkin (teste KMO).

variância relativa ao conjunto de dados original. Em contraste, o método dos fatores principais é usado para identificar as dimensões latentes que apenas mostrem o que as variáveis têm em comum.

Segundo Hair et al. (2005), quando o número de variáveis for maior que 30 ou quando as comunalidades excederem 0,60, os dois métodos chegam a um mesmo resultado.

Não há ainda um método exato para decidir o número de fatores a se extrair, mas os seguintes critérios² são muito utilizados: Critério da raiz latente; Critério *a priori*; Critério de porcentagem de variância e Critério do gráfico *scree*.

A rotação fatorial funciona como uma forma de transformação e consiste em rotacionar os eixos de referências dos fatores, até que uma posição seja alcançada. O efeito final dessa rotação é redistribuir a variância dos fatores para atingir um padrão fatorial mais simples.

Há dois casos de rotação fatorial³, Rotação Ortogonal e Rotação Oblíqua.

A interpretação da matriz de fatores consiste em observar as correlações apresentadas na matriz de cada variável em cada fator. Podemos dizer que quando a carga é alta em um dos fatores, a variável relacionada contribui, e muito, na nomeação desse fator.

Para saber se a carga fatorial é significativa ou não, Hair et al. (2005) propõem os seguintes critérios:

- **Significância prática:** Em síntese, esse critério considera as cargas fatoriais menores que $\pm 0,30$ com o mínimo de significância; $\pm 0,40$ importantes; e maiores que $\pm 0,50$ com significância prática. O valor da carga ao quadrado reflete a variância total da variável explicada pelo fator;

- **Significância estatística:** Nesse critério o conceito de poder do teste e tamanhos da amostra é empregado. A Tabela 2 a seguir revela qual deveria ser o tamanho da amostra para que a carga fatorial seja considerada significativa.

Tabela 2 – Tamanho amostral para referência da significância das cargas fatoriais.

Carga Fatorial	Tamanho da amostra	Carga Fatorial	Tamanho da amostra
0,30	350	0,55	100
0,35	250	0,60	85
0,40	200	0,65	70
0,45	150	0,70	60
0,50	120	0,75	50

Fonte: Cálculos feitos com SOLO Power Analysis, BMDP *Statistical Software*, Inc., 1993.

A validação consiste em avaliar o grau em que os resultados da análise podem ser generalizados para a população. Uma forma de avaliar a possibilidade de generalização é avaliar a repetibilidade dos resultados através de uma amostra. Também pode-se dividir a amostra em duas e comparar os dois modelos fatoriais resultantes.

Há pesquisas nas quais os objetivos envolvem análises posteriores aplicadas aos fatores identificados, e então nesse caso pode-se utilizar uma das opções: selecionar a variável com maior carga fatorial como representativa (variável substituta), substituir o conjunto original por escalas múltiplas ou substituir o conjunto por escores fatoriais.

A técnica de substituir o conjunto por escores fatoriais é caracterizada por considerar todas as cargas fatoriais de determinado fator, mesmo que sejam cargas de variáveis pouco influentes para o fator.

² Maiores informações sobre os critérios para estimação do número de fatores podem ser encontradas em OLIVEIRA (2010).

³ Maiores informações sobre as rotações fatoriais podem ser encontradas em OLIVEIRA (2010).

O escore fatorial representa o grau medido através dos coeficientes referentes ao peso das ponderações de cada variável em cada fator, e podem ser obtidos de diferentes formas tais como: método de mínimos quadrados ponderados, método de regressão e método *ad hoc*.

Análise de Agrupamentos

A análise de agrupamentos é uma técnica multivariada classificada como interdependente, ou seja, todas as variáveis são analisadas ao mesmo tempo com o intuito de encontrar uma estrutura subjacente a todo o conjunto de variáveis. Esta técnica pode ser caracterizada como uma técnica descritiva, sem base teórica e não inferencial.

Embora para a aplicação dessa técnica as suposições de normalidade, homocedasticidade e linearidade não tenham muito peso, é de extrema importância garantir que ao se trabalhar com uma amostra, esta seja representativa da população, e que a multicolinearidade esteja assegurada.

Segundo Mingotti (2005) a análise de agrupamentos, também conhecida como análise de conglomerados, classificação ou *cluster*, tem como objetivo dividir os elementos da amostra, ou população em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis (características) que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características.

A questão inicial que diz respeito à formação desses grupos é a de se saber o quão dois elementos são ou não similares para que pertençam ou não a um mesmo grupo. Para medir tal similaridade, considera-se certa distância (medida de forma métrica) entre os elementos pertencentes à amostra ou população e então procura-se agrupar aqueles que tenham a menor distância.

Dentre as medidas de similaridade e dissimilaridade as mais utilizadas são as seguintes: Distância Euclidiana, Distância generalizada ou

ponderada, Distância de Minkowsky e Distância Euclidiana Média.

A padronização dos dados é sugerida quando se encontra discrepâncias entre escalas e magnitudes de variáveis, pois esses fatos quando não são reflexo de uma relação natural das escalas implicam no uso de medidas de similaridade inadequadas e que tem como consequência agrupamentos que não condizem com a verdadeira semelhança dos dados.

Existem muitas formas para se padronizar os dados. Segundo Hair et al. (2005), a forma mais comum de padronização é a conversão de cada variável em escores padrão (também conhecidos como escores Z) pela subtração da média e divisão pelo desvio-padrão de cada variável.

Os procedimentos para a construção dos agrupamentos são subdivididos em hierárquicos aglomerativos ou divisivos, e não hierárquicos. O primeiro caso é utilizado quando se deseja identificar quantos possíveis *clusters* (ou grupos) os elementos de interesse podem formar, e o segundo é utilizado quando o pesquisador já tem a princípio pré-especificado o número g de grupos. Em se tratando de procedimentos hierárquicos, será focado somente os aglomerativos em razão de serem mais comumente empregados em softwares estatísticos.

Os procedimentos Hierárquicos Aglomerativos partem do princípio de que no início do algoritmo tem-se n clusters, ou seja, cada elemento é considerado como um subgrupo. Nas etapas que se seguem, dois elementos (ou grupos) vão se combinando através de similaridades, reduzindo então o número de agrupamentos em uma unidade a cada passo até que se tenha g grupos.

A ideia de hierarquia é sintetizada no sentido de que se dois elementos são unidos em alguma etapa eles não poderão mais ser separados. Sua representação pode ser visualizada na forma de dendrograma ou gráfico em árvore.

No final do algoritmo todos os elementos pertencem a um único grupo. Por essa razão,

esses procedimentos são também conhecidos como métodos construtivos. Os cinco principais procedimentos hierárquicos⁴ são: Ligação Simples, Ligação completa, Ligação Média, Método do Centróide e Método de Ward.

A questão mais desconfortável para o pesquisador é saber quando se deve interromper o algoritmo de agrupamento, ou seja, quantos grupos devem ser formados. De acordo com Hair et al. (2005), para auxiliar nessa decisão existem os seguintes critérios⁵: Análise do nível de fusão (distância), Análise do nível de similaridade, Análise da soma de quadrados entre os grupos, Estatística Pseudo F, Estatística Pseudo T^2 , Correlação semiparcial ou Método de Ward e Estatística CCC (*Cubic Clustering Criterion*).

Os procedimentos não hierárquicos diferem principalmente dos métodos hierárquicos aglomerativos no sentido de que a princípio já se tem definido um número de clusters.

O algoritmo funciona de forma que a cada passo do processo de agrupamento pode ocorrer a junção ou divisão de um grupo. Isso significa que nem sempre elementos pertencentes a um cluster permanecerão juntos até o final do agrupamento. Sendo assim, não é possível construir o diagrama em árvore (ou dendrograma).

Dois métodos não hierárquicos⁶ que podem ser utilizados são: Método das K-Médias e Método *Fuzzy c-means*.

Ao final do agrupamento deve-se garantir que os grupos formados sejam representativos da população, e assim possam ser generalizados para outros objetos.

Um método que pode auxiliar a decidir se o agrupamento é ou não representativo é a validade

preditiva. Para a execução desse método, seleciona-se uma variável bem definida dentro de padrões teóricos, alocam-se os elementos em grupos tomados como referência o próprio agrupamento e por fim realiza-se algum teste para verificar se esses novos grupos diferem entre si.

Um outro método de agrupamento que é pouco utilizado é o método *two-way joining*, que segundo Rocha et al. (2000), é representado por um gráfico de escala de cores que expressa, na leitura vertical, os diferenciais entre cada variável avaliada pelos especialistas, quão desigual é a avaliação para as diferentes atividades e, na horizontal, uma diferenciação das escalas apontadas dentro de cada atividade. Uma maior variação de cores indica maior heterogeneidade.

Segundo Firetti et al. (2012), com esta técnica, é possível analisar o agrupamento de casos (atividades) e variáveis simultaneamente por meio de suas similaridades e avaliar, principalmente, a situação atual do aparato para geração e transferência de tecnologias frente às necessidades de capacitação e contribuição da pesquisa e desenvolvimento apontados pelos especialistas, para que as atividades identificadas possam de fato expressar seu potencial.

Resultados

Para esse estudo foram coletadas informações sobre os municípios pertencentes à 10ª Região Administrativa (Região Administrativa de Presidente Prudente), em armazéns de dados da Fundação SEADE e do Instituto de Economia Agrícola, o Levantamento de Unidade de Produção Agropecuária (LUPA) do estado de São Paulo, referentes ao ano de 2008. O quadro 1 apresenta

⁴ Maiores informações sobre os procedimentos hierárquicos podem ser encontradas em OLIVEIRA (2010).

⁵ Maiores informações sobre os critérios para definir o número de clusters podem ser encontrados em OLIVEIRA (2010).

⁶ Maiores informações sobre os procedimentos não hierárquicos podem ser obtidas em OLIVEIRA (2010).

as variáveis coletadas no armazém de dados da Fundação SEADE. Para maiores detalhes sobre as variáveis utilizadas consulte OLIVEIRA (2010).

Quadro 1 – Tamanho amostral para referência da significância das cargas fatoriais.

Variável	Descrição
despmunagric.	Total de Despesas Municipais - Agricultura e Organização Agrária.
vlradicfiscagric.	Valor Adicionado Fiscal da Agricultura, Pecuária e Outros Produtos Animais.
crdtragr.	Crédito Rural Agricultura.
vea.	Vínculos Empregaticios na Agropecuária.
partVEA	Participação dos Vínculos Empregaticios na Agropecuária no Total de Vínculos
rendMedioVEA	Rendimento Médio nos Vínculos Empregaticios na Agropecuária

Fonte: Os autores.

O quadro 2 apresenta as variáveis coletadas no Instituto de Economia Agrícola, o Levantamento de Unidade de Produção Agropecuária (LUPA) do estado de São Paulo.

Quadro 2 – Tamanho amostral para referência da significância das cargas fatoriais.

Variável	Descrição
Areatotal	Número total de Unidades de Produção Agropecuária (UPA) do município
areacultperene	Número de Unidades de Produção Agropecuária do município com culturas perenes ⁷
areatemp	Número de Unidades de Produção Agropecuária do município com culturas temporárias ⁸
de0a10	Número de Unidades de Produção Agropecuária do município com (0, 10] ha ⁹
de10a20	Número de Unidades de Produção Agropecuária do município com (10, 20] ha
de20a50	Número de Unidades de Produção Agropecuária do município com (20, 50] ha
mais50	Número de Unidades de Produção Agropecuária do município com mais de 50 ha
GF_TOTalgodao	Área de cultura de algodão cultivada no município em ha
GF_TOTamendoim	Área de cultura de amendoim cultivada no município em ha

⁷ Cultura perene é a cultura que após ser plantada e concluir um ciclo produtivo, não há necessidade de se replantar. (Fonte: <http://www.webartigos.com/articles/24527/1/Cultura-perene/pagina1.html#ixzz1G3BSiLet>)

⁸ Cultura Temporária: São culturas cujo ciclo é de no máximo 1 (um) ano e se caracteriza somente por uma colheita. (<http://www.portaldecontabilidade.com.br/guia/atividaderrural.htm>)

⁹ ha = Hectares

GF_TOTamendoim	Área de cultura de amendoim cultivada no município em ha
GF_TOTarroz	Área de cultura de arroz cultivada no município em ha
GF_TOTaveia	Área de cultura de aveia cultivada no município em ha
GF_TOTcafé	Área de cultura de café cultivada no município em ha
GF_TOTcolza	Área de cultura de colza cultivada no município em ha
GF_TOTervilha	Área de cultura de ervilha cultivada no município em ha
GF_TOTfeijao	Área de cultura de feijão cultivada no município em ha
GF_TOTfeijaooco	Área de cultura de feijão de corda cultivada no mun. em ha
GF_TOTgirassol	Área de cultura de girassol cultivada no município em ha
GF_TOTlentilha	Área de cultura de lentilha cultivada no município em ha
GF_TOTmilho	Área de cultura de milho cultivada no município em ha
GF_TOTmilhodoc	Área de cultura de milho doce cultivada no município em ha
GF_TOTmilhosaf	Área de cultura de milho safrinha cultivada no mun. em ha
GF_TOTsoja	Área de cultura de soja cultivada no município em ha
GF_TOTsorgo	Área de cultura de sorgo cultivada no município em ha
GF_TOTsorgofor	Área de cultura de sorgo forrageiro cultivada no município em ha
GF_TOTsorgovas	Área de cultura de sorgo vassoura cultivada no município em ha
GF_TOTtrigo	Área de cultura de trigo cultivada no município em ha

Fonte: Os autores.

Utilizando-se o Alpha de Cronbach nas 32 variáveis selecionadas, obteve-se o valor de alpha de 0,003 para as variáveis sem transformação e 0,800 para as variáveis padronizadas. O teste KMO, segundo a tabela 2, obteve um valor insuficiente (0,497), indicando a necessidade de eliminação de uma ou mais variáveis para que seja possível a aplicação da análise fatorial.

Na figura 2 apresenta-se o resultado do Alpha de Cronbach e do teste KMO, as variáveis que foram eliminadas da análise estão assinaladas em amarelo.

Figura 2 – Resultado do Alpha de Cronbach e do teste KMO do script do programa R.

```

raw_alpha std.alpha G6(smc) average_r mean sd
0.0027 0.8 1 0.11 290214 769503

Reliability if an item is dropped:
raw_alpha std.alpha G6(smc) average_r
despmunagric 0.0040 0.79 1.00 0.11
vlradicfiscagric -0.0019 0.78 1.00 0.10
vea 0.0027 0.78 1.00 0.10
crdLragr 0.0194 0.80 1.00 0.12
partVEA 0.0027 0.81 1.00 0.12
rendMedioVEA 0.0027 0.80 1.00 0.12
arcatotal 0.0027 0.78 0.94 0.10
areacultperene 0.0027 0.79 1.00 0.11
arcatcmp 0.0027 0.78 1.00 0.10
de0a10 0.0027 0.79 0.95 0.11
de10a20 0.0027 0.79 0.95 0.11
de20a50 0.0027 0.78 0.95 0.11
maIs50 0.0027 0.77 0.95 0.10
GF_TOTalgodao 0.0027 0.80 1.00 0.11
GF_TOTamendoim 0.0027 0.78 1.00 0.10
GF_TOTarroz 0.0027 0.81 1.00 0.12
GF_TOTaveia 0.0027 0.79 1.00 0.11
GF_TOTcafé 0.0027 0.80 1.00 0.12
GF_TOTcolza 0.0027 0.80 1.00 0.12
GF_TOTervilha 0.0027 0.80 1.00 0.12
GF_TOTfeijao 0.0027 0.80 1.00 0.11
GF_TOTfeijaoco 0.0027 0.81 1.00 0.12
GF_TOTgirassol 0.0027 0.81 1.00 0.12
GF_TOTlentilha 0.0027 0.81 1.00 0.12
GF_TOTmilho 0.0027 0.79 1.00 0.11
GF_TOTmilhodoc 0.0027 0.81 1.00 0.12
GF_TOTmilhosaf 0.0027 0.81 1.00 0.12
GF_TOTsoja 0.0027 0.79 1.00 0.11
GF_TOTSorgo 0.0027 0.79 1.00 0.11
GF_TOTSorgofofor 0.0027 0.81 1.00 0.12
GF_TOTSorgovas 0.0027 0.80 1.00 0.12
GF_TOTtrigo 0.0027 0.79 1.00 0.11

kmo.test(dados[,c("despmunagric","vlradicfiscagric","vea","crdLragr",
+ "partVEA","rendMedioVEA","arcatotal","areacultperene","areatemp",
+ "de0a10","de10a20","de20a50","maIs50","GF_TOTalgodao",
+ "GF_TOTamendoim","GF_TOTarroz","GF_TOTaveia","GF_TOTcafé",
+ "GF_TOTcolza","GF_TOTervilha","GF_TOTfeijao","GF_TOTfeijaoco",
+ "GF_TOTgirassol","GF_TOTlentilha","GF_TOTmilho","GF_TOTmilhodoc",
+ "GF_TOTmilhosaf","GF_TOTsoja","GF_TOTSorgo","GF_TOTSorgofofor",
+ "GF_TOTSorgovas","GF_TOTtrigo")])
[1] 0.4968779

```

Fonte: Os autores.

Após a eliminação das variáveis em destaque na figura 2, o valor de Alpha de Cronbach para as variáveis padronizadas aumentaram para 0,85 e o valor do teste KMO aumentou para 0,757, classificando-o como ótimo segundo a tabela 2, vide a figura 3.

Figura 3 – Resultado do Alpha de Cronbach e do teste KMO do script do programa R.

```

raw_alpha std.alpha G6(smc) average_r mean sd
0.85 0.85 0.94 0.30 1.7e-17 0.6

Reliability if an item is dropped:
raw_alpha std.alpha G6(smc) average_r
Z.despmunagric 0.83 0.83 0.91 0.29
Z.vlradicfiscagric 0.83 0.83 0.92 0.28
Z.partVEA 0.83 0.83 0.92 0.28
Z.areacultperene 0.86 0.86 0.94 0.33
Z.areatemp 0.83 0.83 0.93 0.29
Z.GF_TOTamendoim 0.85 0.85 0.93 0.32
Z.GF_TOTarroz 0.85 0.85 0.93 0.32
Z.GF_TOTcafé 0.86 0.86 0.94 0.34
Z.GF_TOTcolza 0.82 0.82 0.92 0.28
Z.GF_TOTfeijao 0.83 0.83 0.92 0.28
Z.GF_TOTmilho 0.86 0.86 0.94 0.34
Z.GF_TOTSorgo 0.83 0.83 0.93 0.30
Z.GF_TOTtrigo 0.84 0.84 0.93 0.31

Item statistics
n r r.cor r.drop mean sd
Z.despmunagric 54 0.71 0.72 0.63 1.9e-17 1
Z.vlradicfiscagric 54 0.79 0.81 0.73 -1.0e-18 1
Z.partVEA 54 0.78 0.78 0.72 -1.8e-17 1
Z.areacultperene 54 0.35 0.28 0.23 2.3e-17 1
Z.areatemp 54 0.75 0.74 0.69 1.6e-17 1
Z.GF_TOTamendoim 54 0.45 0.43 0.34 4.7e-17 1
Z.GF_TOTarroz 54 0.49 0.47 0.38 8.0e-17 1
Z.GF_TOTcafé 54 0.33 0.25 0.21 -7.6e-18 1
Z.GF_TOTcolza 54 0.82 0.84 0.77 -1.0e-17 1
Z.GF_TOTfeijao 54 0.78 0.80 0.72 1.2e-17 1
Z.GF_TOTmilho 54 0.29 0.22 0.17 1.6e-17 1
Z.GF_TOTSorgo 54 0.67 0.64 0.59 -6.7e-18 1
Z.GF_TOTtrigo 54 0.57 0.54 0.47 2.4e-17 1

> kmo.test(ddz)
[1] 0.7572492

```

Fonte: Os autores.

Análise descritiva dos dados

A análise descritiva dos dados por mais que possam parecer comuns e inconsequentes são de extrema importância na utilização de qualquer técnica estatística. Elas podem oferecer uma perspectiva de como se comportam os dados e de como são as inter-relações existentes entre elas. Na Tabela 3, seguem algumas medidas descritivas referentes às variáveis em estudo.

Tabela 3 – Medidas descritivas para as variáveis referentes aos produtos em estudo.

Variáveis	Produto	Média	Variância	Desvio Padrão	Assimetria	Curtose
X1	despmunagric	394370,360	138626748069	372326,100	Positiva	Platicúrtica
X2	viradicfiscagric	3503777,194	3,015802* ¹⁰	5491632,000	Positiva	Leptocúrtica
X3	partVEA	17,315	155,002	12,450	Positiva	Leptocúrtica
X4	areaculperene	138,278	33624,920	183,371	Positiva	Leptocúrtica
X5	areastemp	230,963	35749,700	189,076	Positiva	Platicúrtica
X6	GF_TOTamendoim	168,231	205706,600	453,549	Positiva	Leptocúrtica
X7	GF_TOTarroz	2,143	5,810	2,410	Positiva	Leptocúrtica
X8	GF_TOTcafé	207,023	192435,100	438,674	Positiva	Leptocúrtica
X9	GF_TOTcolza	6,150	2,645	1,626	Positiva	Platicúrtica
X10	GF_TOTfeijao	51,212	5441,722	73,768	Positiva	Leptocúrtica
X11	GF_TOTmilho	599,588	1037166,000	1018,414	Positiva	Leptocúrtica
X12	GF_TOTsoργο	79,705	26937,720	164,127	Positiva	Leptocúrtica
X13	GF_TOTtrigo	26,050	999,045	31,608	Positiva	Platicúrtica

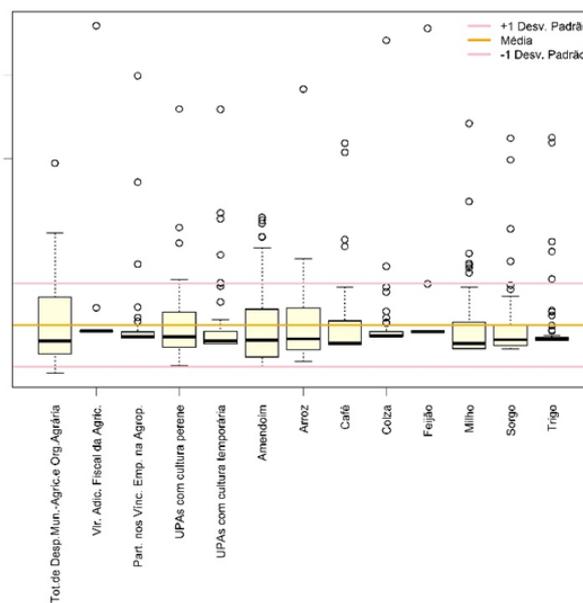
Fonte: Os autores.

Analisando a Tabela 3, observa-se que todas as variáveis possuem assimetria positiva, ou seja, os dados estão concentrados à esquerda da distribuição, próximas do valor zero. As variáveis X1, X2, X6, X8 e X11 possuem variância muito alta indicando que existem municípios com valores muito altos e municípios com valores muito baixos para cada uma dessas variáveis.

Através das medidas de curtose nota-se que as variáveis X1, X5, X9 e X13 apresentam curvatura achatada (Platicúrtica), o que significa que não há um acúmulo grande de observações em um determinado ponto da distribuição de cada uma dessas variáveis. Enquanto que as outras variáveis (X2, X3, X4, X6, X7, X10, X11 e X12), apresentam curvatura alongada (Leptocúrtica), o que indica que há um acúmulo grande de observações em um determinado ponto da distribuição de cada uma dessas variáveis.

Na figura 4, são apresentados em forma de gráficos de caixa os dados¹⁰ das variáveis que compõem esse estudo, evidenciando o que as estatísticas descritivas já haviam indicado através dos resultados de assimetria e curtose, todas as variáveis possuem valores discrepantes, *outliers*.

Figura 4 – Gráficos de caixa das variáveis padronizadas em estudo.

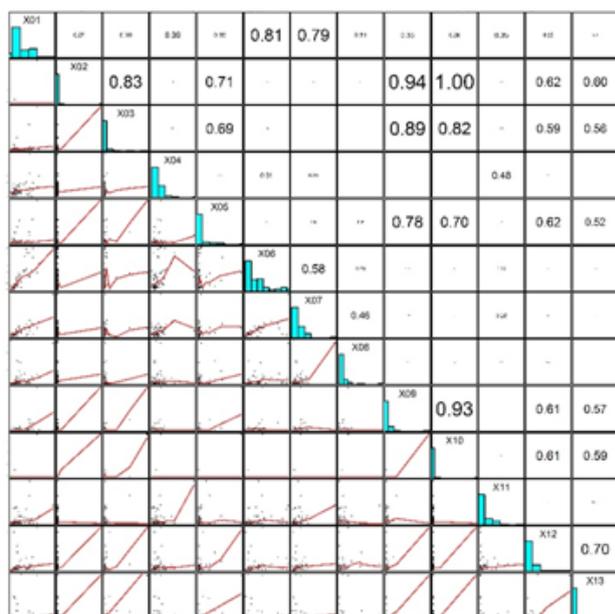


Fonte: Os autores.

Na figura 5 apresenta-se a matriz de gráficos das variáveis referentes à produção de grãos e fibras na 10ª R.A do Estado de São Paulo. Os valores acima da diagonal principal são as correlações bivariadas entre as variáveis em estudo, esses valores se apresentam em tamanhos diferentes representando a intensidade da correlação, ou seja, quanto maior a correlação maior o tamanho da fonte; abaixo da diagonal principal observamos os gráficos de dispersão; e a diagonal principal contém os histogramas que representam a distribuição de cada variável.

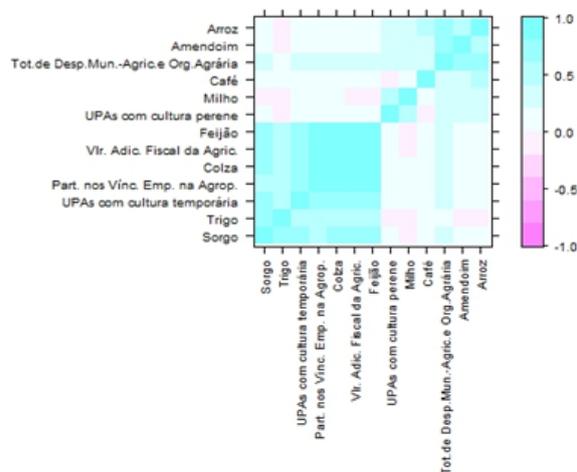
¹⁰ As variáveis estão padronizadas o eixo y apresenta a magnitude dos dados em desvios padrão.

Figura 5 – Matriz de dispersão das variáveis em estudo na 10ª R.A. do Estado de São Paulo.



Uma forma de representar a correlação entre as variáveis é apresentada na figura 6¹¹, utilizando-se da gradação de cores, em que cores mais próximas do ciano significam correlação altamente positiva enquanto que as cores mais próximas de púrpura significam correlação altamente negativa.

Figura 6 – Gráficos de correlação entre as variáveis representada pela gradação de cores.



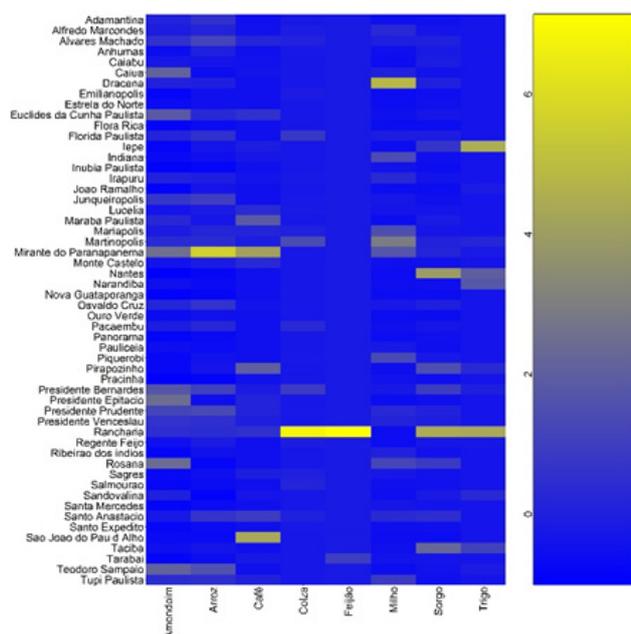
Fonte: Os autores.

Pela figura 7¹² é mostrado um gráfico chamado Two-way Joining, que é uma forma de se avaliar o agrupamento onde a intensidade da cor apresenta em que município, determinada produção está mais relacionada, por exemplo, Rancharia está mais relacionada com as áreas cultivadas com produção de colza e feijão.

¹¹ levelplot(cor[ord, ord], at = do.breaks(c(-1.01, 1.01), 10), xlab="", ylab="", cex=10, scales = list(x = list(rot = 90)))

¹² myImagePlot(dadosstd[,6:13], xLabels=c('Amendoim', 'Arroz', 'Café', 'Colza', 'Feijão', 'Milho', 'Sorgo', 'Trigo'), yLabels = c(as.character(municipio)))

Figura 7 – Gráficos de Two-way joining das variáveis em estudo.



Fonte: Os autores.

Análise Fatorial

A análise fatorial foi aplicada nas 13 variáveis selecionadas através das técnicas indicadas anteriormente e como a técnica sugere, foram verificadas as correlações entre as variáveis, apresentada na figura 5, 33 correlações das 78 correlações calculadas, 42,31%, são iguais ou superiores a que 0,3; que segundo Hair et al. (2005) é um valor indicativo de que a análise fatorial é apropriada.

Resgatando os valores de Alpha de Cronbach de 0,85 e o valor do teste KMO de aproximadamente 0,76, conclui-se então que a técnica análise fatorial pode ser aplicada ao conjunto de dados. Para determinar quantos fatores serão retidos nessa etapa da análise, serão levados em conta três critérios. Primeiramente serão analisados os autovalores maiores que um, ou seja, o será utilizado o critério da raiz latente, vide tabela 4.

Tabela 4 – Autovalores e porcentagem da variância explicada de cada fator

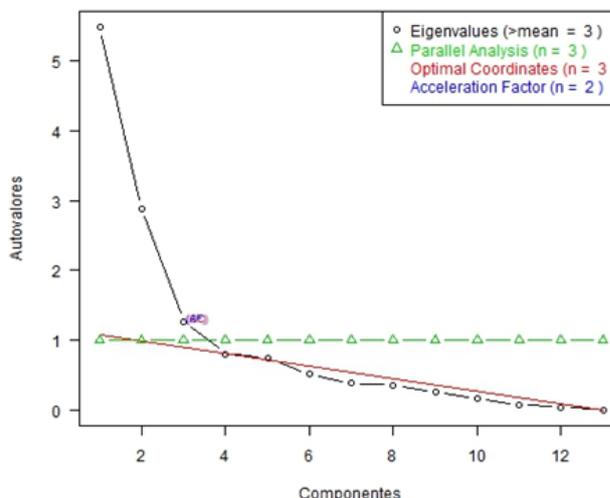
Fatores	Autovalor	Percentual da Variância Explicado pelo Fator	Percentual Acumulado da Variância Explicado pelo Fator
1	5,48884	42,22	42,22
2	2,88514	22,19	64,41
3	1,26732	9,75	74,16
4	0,80605	6,20	80,36
5	0,75104	5,78	86,14
6	0,51288	3,95	90,09
7	0,38264	2,94	93,03
8	0,35651	2,74	95,77
9	0,26028	2,00	97,77
10	0,17313	1,33	99,10
11	0,07711	0,59	99,69
12	0,03651	0,28	99,97
13	0,00254	0,02	99,99

Fonte: Os autores.

Como se pode observar existem três autovalores maiores que um e, portanto três fatores seriam retidos utilizando esse critério. Através da porcentagem da variância explicada acumulada, verifica-se que com três fatores explica-se 74,16% da variância total.

O terceiro critério trata-se da análise visual do gráfico *scree-plot*, apresentado na figura 8, que assim como os dois primeiros critérios analisados, também indica que três fatores são suficientes. E como os três critérios indicam que três fatores são suficientes para serem retidos, portanto, a análise prosseguirá dessa forma.

Figura 8 – Gráfico de scree-plot relacionado a 10ª RA.



Fonte: Os autores.

Segue abaixo, a tabela 5 com as cargas fatoriais presentes em cada fator, segundo variáveis padronizadas.

Tabela 5 – Matriz fatorial sem rotação.

Variáveis	Produto	Fator 1	Fator 2	Fator 3
X1	Total de Desp. Mún. - Agric. e Org. Agrária	0,46	0,82	-0,08
X2	Vlr. Adic. Fiscal da Agric.	0,93	-0,19	0,00
X3	Part. nos Vínc. Emp. na Agrop.	0,88	-0,10	0,13
X4	UPAs com cultura perene	0,14	0,52	0,64
X5	UPAs com cultura temporária	0,83	-0,04	0,02
X6	Area de cultura de Amendoim	0,21	0,79	-0,11
X7	Area de cultura de Arroz	0,22	0,82	-0,30
X8	Area de cultura de Café	0,16	0,41	-0,62
X9	Area de cultura de Colza	0,95	-0,12	0,04
X10	Area de cultura de Feijão	0,92	-0,20	0,00
X11	Area de cultura de Milho	0,05	0,54	0,58
X12	Area de cultura de Sorgo	0,76	-0,13	-0,03
X13	Area de cultura de Trigo	0,70	-0,29	-0,05

Fonte: Os autores.

Verifica-se que as variáveis X4, X8 e X11 apresentam cargas fatoriais muito próximas em mais de um fator, fato que cria dificuldades ao interpretar esses resultados. Para contornar esse problema foi utilizada a rotação fatorial ortogonal Varimax, com o intuito de redistribuir tais cargas.

Segue na tabela 6 como ficou a nova matriz:

Tabela 6 – Matriz fatorial com rotação Varimax.

Variáveis	Produto	Fator 1	Fator 2	Fator 3
X1	Total de Desp. Mún. - Agric. e Org. Agrária	0,25	0,82	0,39
X2	Vlr. Adic. Fiscal da Agric.	0,95	0,05	-0,02
X3	Part. nos Vínc. Emp. na Agrop.	0,89	0,04	0,14
X4	UPAs com cultura perene	0,04	0,13	0,83
X5	UPAs com cultura temporária	0,81	0,14	0,07
X6	Area de cultura de Amendoim	0,01	0,75	0,33
X7	Area de cultura de Arroz	0,01	0,88	0,18
X8	Area de cultura de Café	0,04	0,69	-0,31
X9	Area de cultura de Colza	0,95	0,09	0,05
X10	Area de cultura de Feijão	0,94	0,04	-0,02
X11	Area de cultura de Milho	-0,05	0,16	0,78
X12	Area de cultura de Sorgo	0,76	0,08	-0,03
X13	Area de cultura de Trigo	0,74	-0,06	-0,13

Fonte: Os autores.

Utilizando-se a rotação varimax, houve a redistribuição das cargas fatoriais de forma que os fatores ficassem ortogonais,

Portanto, de acordo com a matriz fatorial com rotação varimax, as variáveis alocam-se da seguinte forma:

Fator 1: X2 (Valor Adicionado Fiscal da Agricultura, Pecuária e Outros Produtos Animais), X3 (Participação dos Vínculos Empregatícios na Agropecuária no Total de Vínculos), X5 (Número de Unidades de Produção Agropecuária do município com culturas temporária), X9 (Área de cultura de colza cultivada no município em ha), X10 (Área de cultura de feijão cultivada no município em ha), X12 (Área de cultura de sorgo cultivada no município em ha) e X13 (Área de cultura de trigo cultivada no município em ha);

Fator 2: X1 (Total de Despesas Municipais - Agricultura e Organização Agrária), X6 (Área de cultura de amendoim cultivada no município em ha), X7 (Área de cultura de arroz cultivada no município em ha) e X8 (Área de cultura de café cultivada no município em ha);

Fator 3: X4 (Número de Unidades de Produção Agropecuária do município com culturas perenes) e X11 (Área de cultura de milho cultivada no município em ha).

Dessa forma, nomearemos o fator 1 como “*PIB agrícola e grãos leguminosos*”, o fator 2 como “*Despesas municipais com agricultura e produção agrícola diversificada*” e fator 3 como “*Agricultura perene*”.

Análise de Agrupamentos

Serão utilizados os descritores (fatores) estatísticos: “*PIB agrícola e grãos leguminosos*” (descritor 1), “*Despesas municipais com agricultura e produção agrícola diversificada*” (descritor 2) e “*Agricultura perene*” (descritor 3). Como referencial teórico, o método hierárquico aglomerativo escolhido para a construção dos

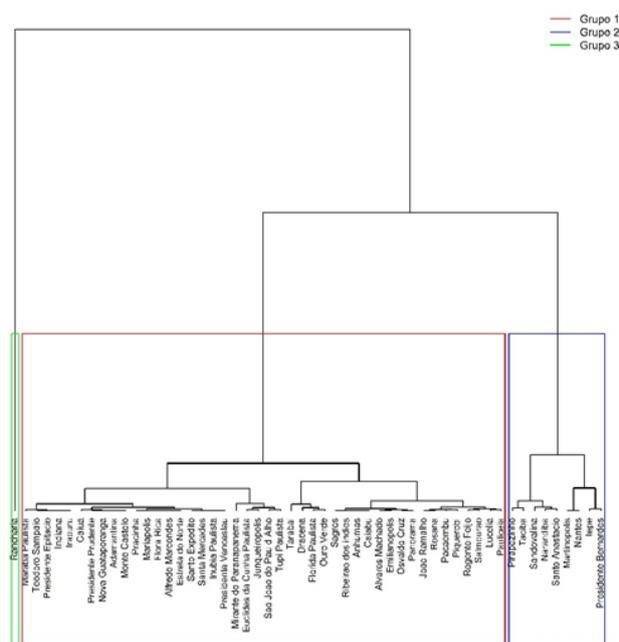
grupos de municípios foi o Método de Ward e como distância foi utilizada a Distância Minkowsky.

Com o intuito de analisar como se comporta a produção de grãos e fibras e de se identificar quais são os núcleos produtivos, no que diz respeito às áreas em produção, os municípios pertencentes a R.A de Presidente Prudente foram alocados segundo os valores de cada descritor em três grupos. A seguir são apresentados os agrupamentos na forma de dendrogramas, cartas temáticas e tabelas com as médias e totais das variáveis de cada descritor.

Análise de agrupamento no descritor: “PIB agrícola e grãos leguminosos”

A figura 9 apresenta o dendrograma do descritor “PIB agrícola e grãos leguminosos”.

Figura 9 - Dendrograma ou árvore hierárquica do agrupamento de municípios da R.A de Presidente Prudente, segundo o descritor “PIB agrícola e grãos leguminosos”.



Fonte: Os autores.

O Grupo 1 é formado por 44 municípios: Adamantina, Alfredo Marcondes, Álvares Machado, Anhumas, Caiabu, Caiuá, Dracena, Emilianópolis, Estrela do Norte, Euclides da Cunha Paulista, Flora Rica, Florida Paulista, Indiana, Inúbia Paulista, Irapuru, João Ramalho, Junqueirópolis, Lucélia, Marabá Paulista, Mariópolis, Mirante do Paranapanema, Monte Castelo, Nova Guataporanga, Osvaldo Cruz, Ouro Verde, Pacaembu, Panorama, Paulicéia, Piquerobi, Pracinha, Presidente Epitácio, Presidente Prudente, Presidente Venceslau, Regente Feijó, Ribeirão dos Índios, Rosana, Sagres, Salmourão, Santa Mercedes, Santo Expedito, São João do Pau d’Alho, Tarabai, Teodoro Sampaio e Tupi Paulista.

O Grupo 2 organizou-se com nove municípios: Iepê, Martinópolis, Nantes, Narandiba, Pirapozinho, Presidente Bernardes, Sandovalina, Santo Anastácio e Taciba.

E o Grupo 3 é formado pelo município de Rancheira.

Para que se pudesse ter uma melhor compreensão da magnitude desses grupos, no que diz respeito ao descritor “PIB agrícola e grãos leguminosos”, as variáveis originais foram retomadas na tabela 7:

Tabela 7 – Área média, área total e número de municípios para cada variável do descritor “PIB agrícola e grãos leguminosos”, em relação aos grupos formados.

Grupo	Descrição	Média	Total	Nº de Municípios
1	Vir. Adic. Fiscal da Agric.	1.321.087,955	29.063.935,000	22
	Part. nos Vinc. Emp. na Agrop.	16,712	735,310	44
	UPAs com cultura temporária	223,705	9.843,000	44
	Área de cultura de Colza	6,150	12,300	2
	Área de cultura de Feijão	51,994	1.715,800	33
	Área de cultura de Sorgo	16,823	218,700	13
	Área de cultura de Trigo	3,700	3,700	1
Total				44
2	Vir. Adic. Fiscal da Agric.	6.787.597,250	54.300.778,000	8
	Part. nos Vinc. Emp. na Agrop.	19,703	177,330	9
	UPAs com cultura temporária	225,000	2025,000	9
	Área de cultura de Colza	-	-	-
	Área de cultura de Feijão	47,529	332,700	7
	Área de cultura de Sorgo	118,883	713,300	6
Área de cultura de Trigo	-	-	-	
Total				9

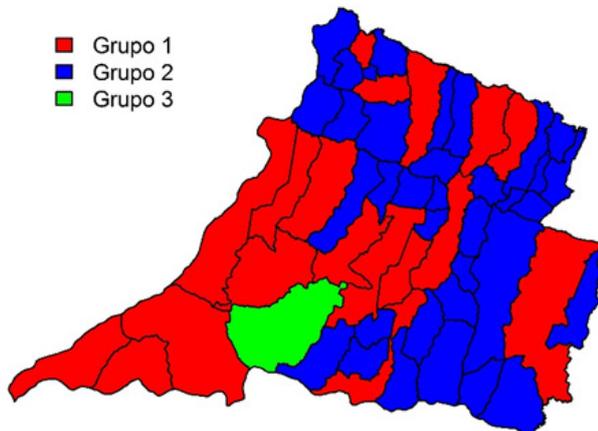
	Vir. Adic. Fiscal da Agric.	25.252.380,000	25.252.380,000	1
	Part. nos Vinc. Emp. na Agrop.	22,370	22,370	1
	UPAs com cultura temporária	604,000	604,000	1
3	Área de cultura de Colza	-	-	-
	Área de cultura de Feijão	-	-	-
	Área de cultura de Sorgo	662,100	662,100	1
	Área de cultura de Trigo	48,400	48,400	1
	Total			1

Fonte: Os autores.

No grupo 1 pode-se notar que todas as variáveis têm valores em pelo menos um município, no grupo 2 apenas as variáveis área de cultura e colza e trigo não estão presentes em pelo menos um município e no grupo 3 as variáveis área de cultura de colza e feijão não estão presentes em pelo menos um município. Para as variáveis que estão presentes em mais de um grupo observa-se que o grupo 1 tem os valores em média menores do que o grupo 2 e o grupo 2 têm valores em média menores do que o grupo 3, com exceção de feijão e sorgo do grupo 2 que é menor que os valores médios do grupo 1.

Para visualização cartográfica dos grupos de municípios formados segundo o descritor “PIB agrícola e grãos leguminosos”, observe a Figura 10.

Figura 12 – Carta temática dos municípios da R.A. de Presidente Prudente, segundo o descritor “Despesas municipais com agricultura e produção agrícola diversificada” no ano de 2008, pela análise fatorial.

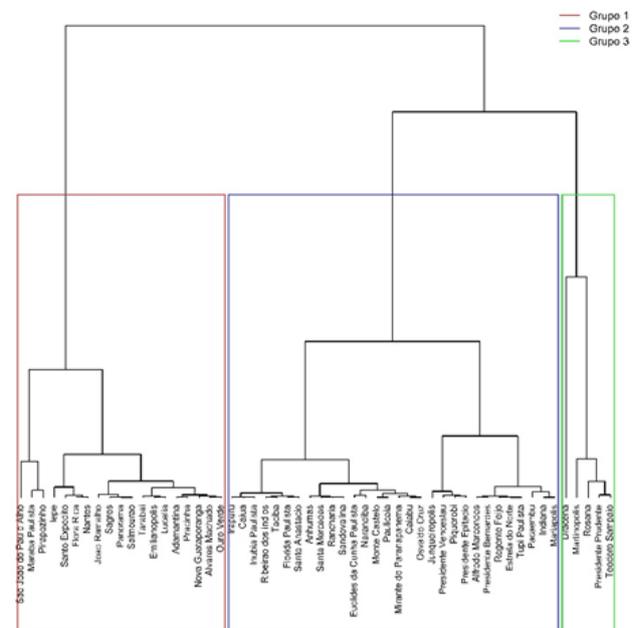


Fonte: Os autores.

Análise de agrupamento no descritor: “Agricultura perene”

A figura 13 apresenta o dendrograma do descritor “Agricultura perene”.

Figura 13 - Dendrograma ou árvore hierárquica do agrupamento de municípios da R.A de Presidente Prudente, segundo o descritor “Agricultura perene”.



Fonte: Os autores.

O Grupo 1 é formado por 19 municípios: Adamantina, Álvares Machado, Emilianópolis, Flora Rica, Iepê, João Ramalho, Lucélia, Marabá Paulista, Nantes, Nova Guataporanga, Ouro Verde, Panorama, Pirapozinho, Pracinha, Sagres, Salmourão, Santo Expedito, São João do Pau d’Alho e Tarabai.

O Grupo 2 organizou-se com 30 municípios: Alfredo Marcondes, Anhumas, Caiabu, Caiuá, Estrela do Norte, Euclides da Cunha Paulista, Florida Paulista, Indiana, Inúbia Paulista, Irapuru, Junqueirópolis, Mariópolis, Mirante do Paranapanema, Monte Castelo, Narandiba, Osvaldo Cruz, Pacaembu, Paulicéia, Piquerobi, Presidente

Bernardes, Presidente Epitácio, Presidente Venceslau, Rancharia, Regente Feijó, Ribeirão dos Índios, Sandovalina, Santa Mercedes, Santo Anastácio, Taciba e Tupi Paulista.

E o Grupo 3 é formado por 5 municípios: Dracena, Martinópolis, Presidente Prudente, Rosana e Teodoro Sampaio.

Para que se pudesse ter uma melhor compreensão da magnitude desses grupos, no que diz respeito ao descritor “Agricultura perene”, as variáveis originais foram retomadas, e então foram obtidos os seguintes resultados na tabela 9:

Tabela 9 – Área média, área total e número de municípios para cada variável do descritor “Agricultura perene”, em relação aos grupos formados.

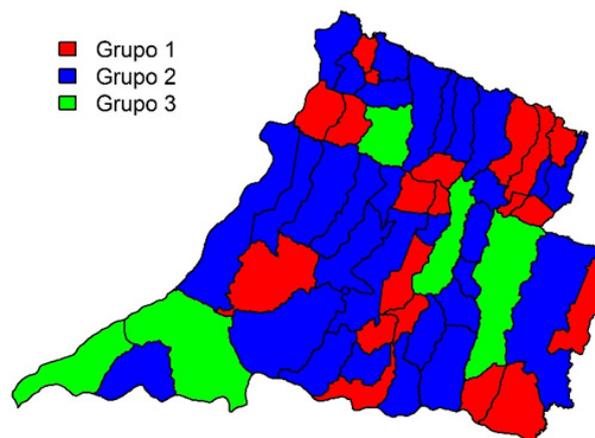
Grupo	Descrição	Média	Total	Nº de Municípios
1	UPAs com cultura perene	55,579	1.056,000	19
	Área de cultura de Milho	602,600	10.244,200	17
	Total			19
2	UPAs com cultura perene	171,900	5.157,000	30
	Área de cultura de Milho	580,103	17.403,100	30
	Total			30
3	UPAs com cultura perene	250,800	1.254,000	5
	Área de cultura de Milho	706,260	3.531,300	5
	Total			5

Fonte: Os autores.

O grupo 1 tem o menor valor médio no número de UPAs com cultura perene e o segundo maior valor médio em área de cultura de milho. O grupo 2 tem o segundo maior valor médio no número de UPAs com cultura perene e o menor valor médio na área de cultura de milho. O grupo 3 tem o maior valor médio para o número de UPAs com cultura perene e a área de cultura de milho.

Para visualização cartográfica dos grupos de municípios formados segundo o descritor “Agricultura perene”, observe a Figura 14.

Figura 14 – Carta temática dos municípios da R.A. de Presidente Prudente, segundo o descritor “Agricultura perene” no ano de 2008, pela análise fatorial.

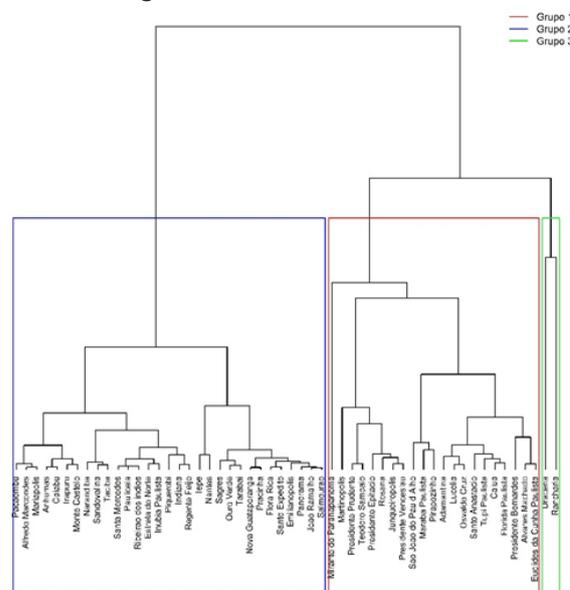


Fonte: Os autores.

Análise de agrupamento nos descritores: “PIB agrícola e grãos leguminosos”, “Despesas municipais com agricultura e produção agrícola diversificada” e “Agricultura perene”

A figura 15 apresenta o dendrograma dos três descritores.

Figura 15 - Dendrograma ou árvore hierárquica do agrupamento de municípios da R.A de Presidente Prudente, segundo os três descritores.



Fonte: Os autores.

O Grupo 1 é formado por 21 municípios: Adamantina, Álvares Machado, Caiuá, Euclides da Cunha Paulista, Florida Paulista, Junqueirópolis, Lucélia, Marabá Paulista, Martinópolis, Mirante do Paranapanema, Osvaldo Cruz, Pirapozinho, Presidente Bernardes, Presidente Epitácio, Presidente Prudente, Presidente Venceslau, Rosana, Santo Anastácio, São João do Pau d'Alho, Teodoro Sampaio e Tupi Paulista.

O Grupo 2 organizou-se com 31 municípios: Alfredo Marcondes, Anhumas, Caiabu, Emilianópolis, Estrela do Norte, Flora Rica, Iepê, Indiana, Inúbia Paulista, Irapuru, João Ramalho, Mariapolis, Monte Castelo, Nantes, Narendiba, Nova Guataporanga, Ouro Verde, Pacaembu, Panorama, Paulicéia, Piquerobi, Pracinha, Regente Feijó, Ribeirão dos Índios, Sagres, Salmourão, Sandovalina, Santa Mercedes, Santo Expedito, Taciba e Tarabai.

E o Grupo 3 é formado por 2 municípios: Dracena e Rancharia.

Para que se pudesse ter uma melhor compreensão da magnitude desses grupos, no que diz respeito dos três descritores, as variáveis originais foram retomadas na tabela 10:

Tabela 10 – Área média, área total e número de municípios para cada variável dos três descritores, em relação aos grupos formados.

Grupo	Descrição	Média	Total	Nº de Municípios
	Total de Desp. Mun. - Agric. e Org. Agrár.	450.112,105	8.552.130,000	19
	Vlr. Adic. Fiscal da Agric.	3.548.218,133	53.223.272,000	15
	Part. nos Vínc. Emp. na Agrop.	14,775	310,230	21
	UPAs com cultura perene	257,286	5.403,000	21
	UPAs com cultura temporária	372,000	7.812,000	21
	Área de cultura de Amendoim	137,875	1.654,500	12
1	Área de cultura de Arroz	1,650	3,300	2
	Área de cultura de Café	314,680	6.293,600	20
	Área de cultura de Colza	7,300	7,300	1
	Área de cultura de Feijão	54,822	986,800	18
	Área de cultura de Milho	602,050	12.041,000	20
	Área de cultura de Sorgo	60,011	540,100	9
	Área de cultura de Trigo	-	-	-
	Total			21

	Total de Desp. Mun. - Agric. e Org. Agrár.	288.213,034	8.358.178,000	29
	Vlr. Adic. Fiscal da Agric.	1.831.429,571	25.640.014,000	14
	Part. nos Vínc. Emp. na Agrop.	19,168	594,220	31
	UPAs com cultura perene	52,871	1.639,000	31
	UPAs com cultura temporária	116,933	3.625,000	31
	Área de cultura de Amendoim	53,293	799,400	15
2	Área de cultura de Arroz	2,340	11,700	5
	Área de cultura de Café	122,627	3.678,800	30
	Área de cultura de Colza	-	-	-
	Área de cultura de Feijão	33,257	698,400	21
	Área de cultura de Milho	448,067	13.442,000	30
	Área de cultura de Sorgo	42,167	379,500	9
	Área de cultura de Trigo	3,700	3,700	1
	Total			31
	Total de Desp. Mun. - Agric. e Org. Agrár.	1.404.105,000	2.808.210,000	2
	Vlr. Adic. Fiscal da Agric.	14.876.903,500	29.753.807,000	2
	Part. nos Vínc. Emp. na Agrop.	15,255	30,510	2
	UPAs com cultura perene	212,500	425,000	2
	UPAs com cultura temporária	517,500	1.035,000	2
	Área de cultura de Amendoim	1.212,400	2.424,800	2
3	Área de cultura de Arroz	-	-	-
	Área de cultura de Café	396,400	792,800	2
	Área de cultura de Colza	5,000	5,000	1
	Área de cultura de Feijão	363,300	363,300	1
	Área de cultura de Milho	2.847,800	5.695,600	2
	Área de cultura de Sorgo	337,250	674,500	2
	Área de cultura de Trigo	48,400	48,400	1
	Total			2

Fonte: Os autores.

O grupo 1 tem o maior valor médio nas variáveis número de UPAs com cultura perene e área de cultura de Colza, o segundo maior valor médio nas variáveis total de despesa municipal com agricultura e organizações agrárias, valor adicionado fiscal da agricultura, pecuária e outros produtos animais, número de UPAs com cultura temporária, área de cultura de amendoim, área de cultura de arroz, área de cultura de café, área de cultura de feijão, área de cultura de milho e área de cultura de sorgo, a variável participação dos vínculos empregatícios na agropecuária no total de vínculos tem o menor valor médio e nenhum município do grupo 1 apresentou área de cultura de trigo.

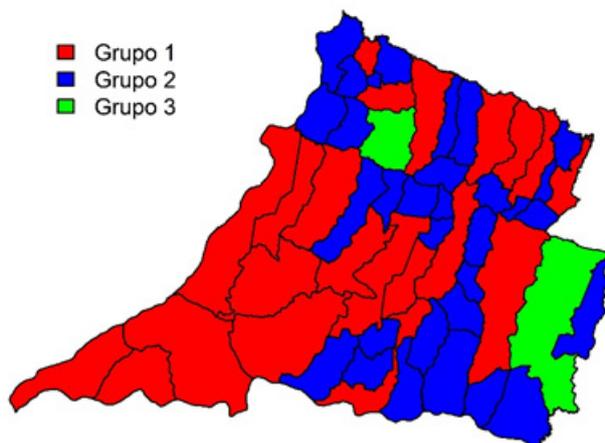
O grupo 2 apresenta o maior valor médio para as variáveis participação dos vínculos empregatícios na agropecuária no total de vínculos e área de cultura de arroz, a variável com o segundo maior valor médio é área de cultura de trigo, e as variáveis com o menor valor médio no grupo 3 são total de despesa municipal com agricultura e organizações agrárias, valor médio nas variáveis valor adicionado

fiscal da agricultura, pecuária e outros produtos animais, número de UPAs com cultura perene, número de UPAs com cultura temporária, área de cultura de amendoim, área de cultura de café, área de cultura de feijão, área de cultura de milho e área de cultura de sorgo e nenhum município do grupo 1 apresentou área de cultura de Colza.

O grupo 3 apresenta o maior valor médio para as variáveis total de despesa municipal com agricultura e organizações agrárias, valor adicionado fiscal da agricultura, pecuária e outros produtos animais, número de UPAs com cultura temporária, área de cultura de amendoim, área de cultura de café, área de cultura de feijão, área de cultura de milho, área de cultura de sorgo e área de cultura de trigo, as variáveis com o segundo maiores valores médios são participação dos vínculos empregatícios na agropecuária no total de vínculos, número de UPAs com cultura perene e área de cultura de Colza, e nenhum município do grupo 3 apresentou área de cultura de arroz.

Para visualização cartográfica dos grupos de municípios formados segundo os três descritores, observe a Figura 16.

Figura 16 – Carta temática dos municípios da R.A. de Presidente Prudente, segundo os três descritores no ano de 2008, pela análise fatorial.



Fonte: Os autores.

Conclusões

O presente estudo retratou as características da produção de grãos e fibras na R.A. de Presidente Prudente no ano de 2008, para dar embasamento a futuras discussões e trabalhos sobre o tema.

Com a aplicação da técnica Análise Fatorial, resumiu-se o conjunto de 13 variáveis para três descritores estatísticos (fatores): “PIB agrícola e grãos leguminosos”, “Despesas municipais com agricultura e produção agrícola diversificada” e “Agricultura perene”.

Os agrupamentos de municípios possibilitam avaliar a existência ou não de núcleos produtivos locais, por descritor ou no conjunto de descritores, de forma a orientar instituições públicas e privadas do Estado de São Paulo na reflexão sobre como se encontra o atual cenário de produção de grãos e fibras da região estudada, e então tomar decisões que possibilitem o desenvolvimento regional.

Apesar da Região Administrativa de Presidente Prudente poder usufruir de um sistema de distribuição trimodal, o que poucas regiões do estado têm à disposição, isso não se traduz na disposição de se produzir em grande escala. Assim sendo a descrição estatística apresentada nesse trabalho motiva e é motivo do descaso político no desenvolvimento e aperfeiçoamento das rodovias, ferrovias e porto fluvial.

Referências

- BARAT, J.; VIDIGAL, A. A. F. V.; GANDRA, M.; DUPAS, G. *Logística e transporte no processo de globalização: oportunidades para o Brasil*. São Paulo: UNESP, 2007. 255 p.
- BARROSO, L. P.; ARTES, R. *Minicurso: Análise Multivariada*. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA (10º SEAGRO e 48º RBRAS), 2003. Lavras: UFLA, 2003. 156 p.

BRASIL. Ministério dos Transportes. *Banco de Informações e Mapas de Transportes - BIT*. Disponível em: <<http://www.transportes.gov.br/bit/estados/port/sp.htm>>. Acesso em: 25 nov. 2010.

Recebido em 29 Abril, 2013 – Received on April 29, 2013
Aceito em 4 Julho, 2014 – Accepted on July 4, 2014

FIRETTI, R.; CAPANEMA, L. M.; FACHINI, C.; TURCO, P. H. N.; VEIGA FILHO, A. A. Análise de variáveis estratégicas para o desenvolvimento da agropecuária da região do pontal do Paranapanema. *Revista de economia e sociologia rural*, Brasília, v. 50, n. 1, p. 141-156, 2012.

HAIR, J. F.; TATHAM, R. L.; ANDERSON, R. E.; BLACK, W. C. *Análise multivariada de dados*. Tradução de Adonai Schlup Sant'Anna; Anselmo Chaves Neto. 5. ed. Porto Alegre: Bookman, 2005.

MICHELETTO, M. *Crescimento da área plantada de grãos e fibras nos últimos 20 anos*. Paraná. 2009. Disponível em: <http://www.deputadomoacirmicheletto.com.br/attachments/265_Crescimento_da_area_plantada.pdf>. Acesso em: 25 out. 2010.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: UFMG, 2005. 297 p.

OLIVEIRA, E. C. *Identificação e elaboração de variáveis estatísticas no estudo da produção de grãos e fibras, na região administrativa de Presidente Prudente*. 2010. 77 p. Monografia (Especialização em Estatística) - Universidade Estadual de Londrina, Londrina, 2010.

ROCHA, A. D.; OKABE, I.; MARTINS, M. E. A; MACHADO, P. H. B; MELLO, T. C. Qualidade de vida, ponto de partida ou resultado final?. *Ciência e saúde coletiva*, Rio de Janeiro, v. 5, n. 1, p. 61 – p. 83. 2000.

SÃO PAULO (Estado). Fundação Sistema Estadual de Análise de Dados (SEADE). *Região Administrativa de Presidente Prudente*. São Paulo, 2007. Disponível em: <http://www.emplo.sp.gov.br/outros/caravana_trabalho/boletim_presidentepudente.pdf>. Acesso em: 25 out. 2010.

Anexo

Segue abaixo as rotinas utilizadas para a obtenção dos gráficos¹³.

```
##### Gráfico de caixas (Box-plot)
jpeg(filename = 'figboxplot.jpg', width=500, height=500, pointsize=12, quality=100,
bg='white')
par(mar=c(13,1.5,.5,.5), cex=1, lwd=.8)
boxplot(dadosstd, whiskly=3, cex=1, cex.axis=8, las=2, col='lightyellow', names=c('Tot.
de Desp.Mun.-Agric.e Org.Agrária', 'Vlr. Adic. Fiscal da Agric.', 'Part. nos Vnc. Emp. na
Agrop.', 'UPAs com cultura perene', 'UPAs com cultura temporária', 'Amendoim', 'Arroz', 'C
afê', 'Colza', 'Feijão', 'Milho', 'Sorgo', 'Trigo'))
xi <- 0 + seq(length(xb))
abline(h=1, lwd=18, col = "pink")
abline(h=0, lwd=18, col = "orange")
abline(h=-1, lwd=18, col = "pink")
legend('topright', lwd=18, c[+1 Desv. Padrão', 'Média', '-1 Desv. Padrão'),
col=c('pink', 'orange', 'pink'), text.col='black', lty=1, cex=.8, bty = 'n')
dev.off()
```

```
##### Correlação por gradação de cores
jpeg(filename='figpairs.jpg', width=500, height=500, pointsize = 12, quality = 100, bg =
'white')
par(lwd=.8)
pairs(dadosstd, labels=c('X01','X02','X03','X04','X05','X06','X07','X08','X09','X10','X
11','X12','X13'), cex.labels=10, diag.panel=panel.hist, lower.panel = panel.smooth, upper.
panel = panel.cor)
dev.off()
```

```
##### Correlação por gradação de cores
jpeg(filename = 'figlevelplot.jpg', pointsize = 12, quality = 100, bg = 'white')
par(mar=c(0,0,2.5,2), mai=c(0,5,2,0), lwd=.8)
levelplot(cor[ord, ord], at = do.breaks(c(-1.01, 1.01), 10),
xlab="", ylab="", cex=1, scales = list(x = list(rot = 90)))
dev.off()
```

```
##### Two-Way Joining
jpeg(filename='figTWPJplot.jpg', width=500, height=500, pointsize=12, quality=100,
bg='white')
par(mai=c(7,16,1,1), cex.axis=.7, lwd=.8)
myImagePlot(dadosstd[,6:13], xLabels=c('Amendoim', 'Arroz', 'Cafê', 'Colza', 'Feijão', 'Milho
', 'Sorgo', 'Trigo'), yLabels=c(as.character(municipio)))
dev.off()
```

```
##### Scree-plot
jpeg(filename='figscreeplot.jpg', width=480, height=385, pointsize=12, quality=100,
bg='white')
par(mar=c(4.3,4,3,3), cex=2, mex=2, cex.axis=1, cex.lab=1, cex.sub=2, lwd=1.5, las=1)
plotnScree(nS, main="", ylab='Autovalores', xlab='Componentes')
dev.off()
```

```
##### Dendrogramas
jpeg(filename = 'figded0.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg =
'white')
HClust.0 <- hclust(dist(model.matrix(~1 + RC1 + RC2 + RC3, dda), method="minkowski"),
method="ward")
par(mar=c(0,0,.5,1), cex=1, lwd=.8)
plot(HClust.0, hang=-1, las=2, main="", xlab="", ylab="", axes=F, labels=dda[,2], cex=.6,
sub="")
summary(as.factor(cutree(HClust.0, k = 3))) # Cluster Sizes
legend('topright', legend=c("Grupo 1","Grupo 2","Grupo 3"), cex = .7, lty=1, lwd=8,
col=c("red","blue","green"), bty="n")
by(model.matrix(~1 + RC1 + RC2 + RC3, dda), as.factor(cutree(HClust.0, k = 3)), mean)
rect.hclust(HClust.0, k=3, border=c("blue","red","green"))
dda$hclus0 <- assignCluster(model.matrix(~1 + RC1 + RC2 + RC3, dda), dda, cutree(HClust.0,
k = 3))
dev.off()
```

```
jpeg(filename = 'figded1.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg =
'white')
HClust.1 <- hclust(dist(model.matrix(~1 + RC1, dda), method="minkowski"),
method="ward")
par(mar=c(0,0,.5,1), cex=1, lwd=.8)
plot(HClust.1, hang=-1, las=2, main="", xlab="", ylab="", axes=F, labels=dda[,2], cex=.6,
sub="")
summary(as.factor(cutree(HClust.1, k = 3))) # Cluster Sizes
legend('topright', legend=c("Grupo 1","Grupo 2","Grupo 3"), cex = .7, lty=1, lwd=8,
col=c("red","blue","green"), bty="n")
by(model.matrix(~1 + RC1, dda), as.factor(cutree(HClust.1, k = 3)), mean)
rect.hclust(HClust.1, k=3, border=c("green","red","blue"))
dda$hclus1 <- assignCluster(model.matrix(~1 + RC1, dda), dda, cutree(HClust.1, k = 3))
```

dev.off()

```
jpeg(filename = 'figded2.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg =
'white')
HClust.2 <- hclust(dist(model.matrix(~1 + RC2, dda), method="minkowski"),
method="ward")
par(mar=c(0,0,.5,1), cex=1, lwd=.8)
plot(HClust.2, hang=-1, las=2, main="", xlab="", ylab="", axes=F, labels=dda[,2], cex=.6,
sub="")
summary(as.factor(cutree(HClust.2, k = 3))) # Cluster Sizes
legend('topright', legend=c("Grupo 1","Grupo 2","Grupo 3"), cex = .7, lty=1, lwd=8,
col=c("red","blue","green"), bty="n")
by(model.matrix(~1 + RC2, dda), as.factor(cutree(HClust.2, k = 3)), mean)
rect.hclust(HClust.2, k=3, border=c("blue","green","red"))
dda$hclus2 <- assignCluster(model.matrix(~1 + RC2, dda), dda, cutree(HClust.2, k = 3))
dev.off()
```

```
jpeg(filename = 'figded3.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg =
'white')
HClust.3 <- hclust(dist(model.matrix(~1 + RC3, dda), method="minkowski"),
method="ward")
par(mar=c(0,0,.5,1), cex=1, lwd=.8)
plot(HClust.3, hang=-1, las=2, main="", xlab="", ylab="", axes=F, labels=dda[,2], cex=.6,
sub="")
summary(as.factor(cutree(HClust.3, k = 3))) # Cluster Sizes
legend('topright', legend=c("Grupo 1","Grupo 2","Grupo 3"), cex = .7, lty=1, lwd=8,
col=c("red","blue","green"), bty="n")
by(model.matrix(~1 + RC3, dda), as.factor(cutree(HClust.3, k = 3)), mean)
rect.hclust(HClust.3, k=3, border=c("red","blue","green"))
dda$hclus3 <- assignCluster(model.matrix(~1 + RC3, dda), dda, cutree(HClust.3, k = 3))
dev.off()
```

```
##### Mapas
sp <- readShapePoly("d://Mapas/sao_paulo/sao_paulo.shp")
```

```
pp = sp[sp$RA%in%(10,1),
cores=c("red","blue","green")
grupos=c("Grupo 1","Grupo 2","Grupo 3")
```

```
jpeg(filename = 'F1F2F3.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg =
'white')
par(lwd=.8)
plot(pp)
legend(-53, -21.1, bty = "n", fill = cores, cex = 9, legend = grupos)
t=length(levels(hclus0))
for (g in 1:t) {
pp1=pp[pp$CODMUNIC %in% dda[hclus0==g,1],]
plot(pp1, col=cores[g], add=TRUE)
}
plot(pp, add=TRUE)
dev.off()
```

```
jpeg(filename = 'F1.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg = 'white')
par(lwd=.8)
plot(pp)
legend(-53, -21.1, bty = "n", fill = cores, cex = 9, legend = c("Grupo 1","Grupo 2","Grupo 3"))
t=length(levels(hclus1))
for (g in 1:t) {
pp1=pp[pp$CODMUNIC %in% dda[hclus1==g,1],]
plot(pp1, col=cores[g], add=TRUE)
}
plot(pp, add=TRUE)
dev.off()
```

```
jpeg(filename = 'F2.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg = 'white')
par(lwd=.8)
plot(pp)
legend(-53, -21.1, bty = "n", fill = cores, cex = 9, legend = c("Grupo 1","Grupo 2","Grupo 3"))
t=length(levels(hclus2))
for (g in 1:t) {
pp1=pp[pp$CODMUNIC %in% dda[hclus2==g,1],]
plot(pp1, col=cores[g], add=TRUE)
}
plot(pp, add=TRUE)
dev.off()
```

```
jpeg(filename = 'F3.jpg', width = 500, height = 500, pointsize = 12, quality = 100, bg = 'white')
par(lwd=.8)
plot(pp)
legend(-53, -21.1, bty = "n", fill = cores, cex = 9, legend = c("Grupo 1","Grupo 2","Grupo 3"))
t=length(levels(hclus3))
for (g in 1:t) {
pp1=pp[pp$CODMUNIC %in% dda[hclus3==g,1],]
plot(pp1, col=cores[g], add=TRUE)
}
plot(pp, add=TRUE)
dev.off()
```

¹³ Script completo do programa R pode ser encontrado OLIVEIRA (2010).

