

# ANÁLISE MULTIVARIADA DE DADOS: O CASO DOS COREDES AGRÍCOLAS DO RIO GRANDE DO SUL E DOS MUNICÍPIOS DE SANTA CATARINA

Rafael Pentiado Poerschke\*

**Resumo:** Esta pesquisa sustenta-se na tese da presença de uma modernização conservadora no setor agropecuário brasileiro. Sob a ótica dos COREDES do Rio Grande do Sul e dos municípios de Santa Catarina, infere-se sobre indícios de correlação entre a orizicultura e a concentração da modernização no campo. Para tanto, recorre-se à estatística multivariada, com o intuito de identificar níveis de modernização da agropecuária nos municípios selecionados de 1995 a 1996. No que diz respeito ao Rio Grande do Sul, verificou-se uma forte concentração da agropecuária moderna em uma reduzida parcela dos municípios dos COREDES, em especial, no Conselho da Fronteira Oeste. Para Santa Catarina, em função do elevado número de municípios da amostra, a metodologia não foi consistente para extrair resultados mais pormenorizados sobre concentração. Entretanto, verificou-se a reprodução dos padrões do COREDE Fronteira Oeste em um dos grupos formados por municípios catarinenses. Esses dois grupos são fortemente relacionados ao conjunto de variáveis ligadas à modernização, além de voltados para cultura do arroz.

**Palavras-chave:** Agricultura, Análise multivariada, Modernização.

**Abstract:** This research rests on the thesis of the presence of a conservative modernization in the Brazilian agricultural sector. From the perspective of COREDES (Regional Councils of Development) at Rio Grande do Sul and municipalities of Santa Catarina, it appears on evidence of correlation between the rice production and the concentration of modernization in the field. For this, we resort to multivariate statistics, in order to identify levels of modernization of agriculture in selected municipalities from 1995 to 1996. With regard to Rio Grande do Sul, there was a high concentration of modern agriculture on a small portion of the municipalities of COREDES, especially in West Border Council. In Santa Catarina, due to the high number of municipalities in the sample, the methodology was not consistent to extract results more detailed on concentration. However, it was verified the reproduction the patterns of COREDE West Border in one of the groups formed by the municipalities of Santa Catarina. These two groups are closely related to the set of variables linked to modernization, and focused to rice production.

**Keywords:** Agriculture, Multivariate analysis, Modernization.

## Introdução

A agricultura brasileira sofreu diversas mudanças estruturais e conjunturais nos últimos cinquenta anos. Os meios de informação e transporte passaram por igual revolução. Da agricultura tradicional e pastoreio às colheitadeiras guiadas via Global Positioning System (GPS); das caravelas a navios mercantes precisos que interligam continentes. Mas toda essa panaceia capitalista se deve à Terceira Revolução Industrial, a partir da invenção do primeiro computador. A difusão da tecnologia reduziu as distâncias e, gradualmente, aumentou o inter-relacionamento e a interdependência no espaço econômico. A agricultura, dentro desse desenvolvimento, (re) encontrou sua vocação guiada por inversões ao longo de todo o processo de modernização deflagrado pelo Sistema Nacional de Crédito Rural (SNCR). Contudo, esse processo não pode ser encarado como homogêneo, aceitando a hipótese de

---

\*Mestrando em Economia pela Universidade do Vale do Rio dos Sinos.

efeitos multifacetados, pois ocorre em lugares variados, com diferentes níveis intensidade, além de ser concentrador e/ou excludente.

A agricultura, durante o processo de seu desenvolvimento, seria integrada e responsável pela transferência de mão-de-obra para o setor urbano-industrial; continuaria a alimentar a população; gerar divisas com o comércio do excedente, além de transferir poupança e consumir bens industriais. Esses pontos abrangem a essência das funções da agricultura no desenvolvimento econômico de um país.

No Brasil, o aumento de produção até meados dos anos sessenta esteve baseado em aumento da área plantada. Sob o governo militar, o padrão foi reestruturado com a adoção de uma nova estratégia de modernização refletida sobre o setor agrícola na forma de complexos agroindustriais, reduzindo a agricultura a um subsetor da indústria. De um lado se encontra a indústria, fornecedora de insumos modernos e de maquinário, de outro, as indústrias processadoras de alimentos e matérias-primas agrícolas.

Em uma tentativa de mapear uma “mancha de modernização no estado do Rio Grande do Sul (RS)” focando os Conselhos Regionais de Desenvolvimentos (COREDEs), verificou-se que a técnica de análise multivariada era de fato adequada, gerando resultados plausíveis para o estado. Por meio da ferramenta de análise multivariada aplicada nos dados do Censo Agropecuário 1995/96 agrupou-se casos segundo o nível de similaridade, conforme variáveis que se relacionam com a questão da agricultura moderna, definida em Silva (1999). Esses estudos preliminares apontaram a concentração de tecnologia moderna na agricultura do Rio Grande do Sul<sup>1</sup>, sobretudo, em um compacto grupo de municípios do RS. Esse cross-section forneceu a possibilidade de inferir sobre o grau da modernização agrícola dos Conselhos Regionais de Desenvolvimentos (COREDEs) – predominantemente agrícolas – e em todos os municípios de Santa Catarina (SC).

Sob o olhar de semelhante metodologia e base de dados, no plano dos principais municípios produtores de arroz no Rio Grande do Sul e Santa Catarina, inferiu-se sobre a ocorrência desse fenômeno no estado de catarinense. Esses dois estados alternam ao longo do tempo a liderança da produtividade/rendimento por hectare. Um forte indício de que houve inversões em pesquisa e implementos a fim de elevar os volumes produzidos no mesmo hectare. De todo modo, vale ressaltar que em Santa Catarina o método de plantio é diferente, pois muito se utiliza a técnica do arroz pré-germinado, plantado manualmente nas várzeas. Essa técnica, peculiar da região sudeste asiática se mostra igualmente produtiva lá e no Sul do Brasil. Nesse sentido, sendo ou não esse método predominante na região, a hipótese básica a ser adotada seria de que esses municípios, embora sabidamente produtivos, não se relacionem com a agricultura moderna.

Essa caracterização do método, mão-de-obra intensiva, aponta para uma agricultura mais tradicional, e dependendo do ponto de vista, da visão socioambiental, menos agressiva e, portanto, sustentável. Contudo, a segunda linha de ação seria então essa caracterização do método de produção predominante em Santa Catarina. Havendo indícios de característica de Agricultura Familiar, refinar a pesquisa para buscar a relação da agricultura moderna ou familiar no estado com a cultura do arroz e a forte produtividade alcançada no estado com um método de plantio diferenciado e intensivo em trabalho.

Ainda, com sinais de diferenças entre o estilo de agricultura, evidenciada previamente pelo método de plantio pré-germinado, pode apontar para um outro paradigma da agricultura dentro do Brasil em uma cultura. Contudo, ainda seria preciso ver se esses agricultores se reconhecem como camponeses ou algo que se possa relacionar ao método de agricultura sustentável segundo a caracterização de Caporal e Costabeber (2004).

Esta pesquisa está organizada em quatro seções, sendo a primeira constituída pela presente introdução. A segunda seção apresenta o marco teórico, fonte de dados e método utilizado. A terceira reúne os resultados da pesquisa, e, finalmente, a quarta seção apresenta conclusão do estudo.

### **Marco teórico, fonte de dados e método**

O atual estudo parte de resultados parciais já apresentados com dados de Poerschke et al. (2008). Os mesmos apontaram para uma forte concentração da modernização da agricultura em uma determinada área abrangida por poucos municípios. Municípios esses que se mostraram no período de 1995/96 os principais produtores de arroz do estado do Rio Grande do Sul. Partindo dessa premissa, além de revisar o trabalho efetuado pelos autores, traçou-se uma aproximação desse agrupamento gaúcho com os dados coletados para os municípios de Santa Catarina. Buscou-se identificar a ocorrência desse mesmo padrão ou não de concentração da modernização nos municípios catarinenses, atentando para aqueles intensivos no plantio de arroz, na medida em que o estado catarinense é o segundo maior produtor do Brasil.

Segundo dados da Embrapa (2009), a cultura de arroz do tipo irrigado predomina no Brasil. Até meados da década de 1980, o arroz de terras altas, ou de sequeiro, prevalecia em área e quantidade produzida sobre o método de irrigação por lâmina d'água. O arroz longo (sequeiro) tem sua produção concentrada no Centro-Oeste, enquanto na Região Sul o plantio é predominantemente irrigado. Por razões históricas e naturais, a Região Sul assume uma vantagem e atrai pesquisa, que aplicada vem proporcionando incrementos produtivos com uma estabilidade da área utilizada (POERSCHKE, 2008).

Juntos, Santa Catarina e Rio Grande do Sul respondem por 70% das 12 milhões de toneladas produzidas e 43% da extensão coberta por arroz (CONAB, 2009). Os dois estados, em 2008/09<sup>2</sup>, são também aqueles com o maior índice de produtividade na referida safra. Sendo essa produção fortemente concentrada no estado (CONAB, 2009), como é o caso de Uruguaiana, responde por considerável volume da produção nacional de arroz. Segundo a CONAB (2009), o método irrigado vem se mostrando mais eficiente ao sequeiro e já concentra cerca de 63% da produção brasileira, com apenas 26% da área total cultivada com arroz. Embora a produção catarinense represente um sétimo da gaúcha, a produtividade média em Santa Catarina é ligeira maior, salientando-se o método diferenciado de sistema pré-germinado.

Poerschke et al. (2008) inferiram sobre a distribuição da agricultura moderna, segundo a definição de Silva (1999), nos COREDEs agrícolas, apontou forte concentração de capital e tecnologia. A análise de agrupamento<sup>3</sup> formou um cluster compacto, batizado de grupo I, com apenas 5,76% da amostra, ou seja, oito municípios em um total de 139, sendo sete do COREDE Fronteira Oeste<sup>4</sup> – Alegrete, Itaqui, Rosário do Sul, Santana do Livramento, São Borja, São Gabriel, Uruguaiana; e; Cachoeira do Sul do COREDE Jacuí Centro<sup>5</sup>. Outros três grupos foram encontrados pelo agrupamento, porém, menos intensivos em fatores modernos de produção.

Na análise fatorial das 17 variáveis utilizadas no estudo, associou-se em um fator o número de tratores (TRATOR – 0,628), o valor das receitas advindas de produtos vegetais e animais (RECVEG – 0,736 e RECANI – 0,903), o valor dos financiamentos (FINANC – 0,520), número de máquinas para colheita (MAQCOL – 0,506), o valor da produção vegetal e animal (VALVEG – 0,741 e VALANI – 0,945), a área explorada nos municípios (AREXPL – 0,963), efetivo bovino (GADO – 0,970), a PEA (0,508) e, por fim, a variável que mede

irrigação (IRRIGA – 0,757). O fator 1 tem uma relação muito forte com as variáveis ligadas ao uso intensivo da terra, como GADO, AREXPL, RECANI, VALANI, IRRI e RECVEG. Sem dúvida, além de medir a pecuária de corte, o fator mede a modernização da agricultura, indicando uma forte relação com os insumos modernos, denotando um caráter capital-intensivo. Dessa forma, os autores batizaram o fator 1 como o fator que mede a “agricultura moderna”, fortemente ligada aos insumos tecnológicos e à pecuária.

Segundo as conclusões finais, Poerschke et al. (2008) apontaram o grupo I formado por municípios intensivos na produção de arroz e na pecuária, sendo o único grupo com relação positiva ao fator “agricultura moderna”. Nesse sentido, variáveis como GADO, RECANI, VALANI e IRRI estão relacionadas com a cultura do arroz e o referido grupo I. Essa correlação ocorre na medida em que o arroz irrigado cresce em áreas de várzea irrigado por bombas a fim de formar a lâmina d’água, além de utilizar os implementos para colheita e plantio, no qual a única opção para rotação se restringe a pecuária (EMBRAPA, 2009). Finalmente, com os demais grupos mostraram-se menos dinâmicos, grupos 2 e 3, e um único grupo, de 55 municípios se mostrou relacionado com uma agricultura em maior atraso tecnológico e mais mão-de-obra intensiva.

Segundo o IBGE (2009), os municípios do grupo I produziram 31% do total de arroz em casca produzido no Rio Grande do Sul, bem como 14% do que foi produzido no Brasil em 1995/96. Os mesmos figuraram entre os 10 primeiros em volume de produção de arroz em 1995/96 e se mantiveram presentes no ranking até o Censo de 2006, conforme os dados da Tabela 1.

**Tabela 1** - Produção de arroz em casca – municípios selecionados do Rio Grande do Sul, comparativo 1995 e 2006 (toneladas), arroz (base casca)

Arroz (base casca)					
Município	1995	(%)	Município	2006	(%)
Uruguaiana (G I)	535.680	10,63	Uruguaiana (G I)	590.329	7,33
Sta Vitória do Palmar	423.000	8,39	Itaqui (G I)	458.118	7,23
Itaqui (G I)	355.000	7,04	Alegrete (G I)	390.000	6,06
Arroio Grande	224.250	4,45	Dom Pedrito	335.835	5,43
Alegrete (G I)	221.015	4,39	Sta Vitória do Palmar	324.619	4,91
São Borja (G I)	210.000	4,17	São Borja	314.069	4,74
Dom Pedrito	196.650	3,90	Cachoeira do Sul (G I)	248.899	4,50
Mostardas	156.510	3,11	Arroio Grande	243.750	3,33
Cachoeira do Sul (G I)	151.780	3,01	São Gabriel (G I)	210.925	3,29
São Gabriel (G I)	150.000	2,98	Mostardas	196.446	3,20
Rosário do Sul (G I)	101.465	2,01	Rosário do Sul (G I)	138.026	3,29
Santana do Livramento (G I)	52.800	1,05	Santana do Livramento (G I)	75.020	3,20
Rio Grande do Sul	5.040.104		Rio Grande do Sul	6.786.242	
Brasil	11.226.064		Brasil	11.526.685	

Fonte: IBGE (2009).

A mesma metodologia aplicada em Poerschke et al. (2008) foi utilizada primeiramente na totalidade dos municípios de Santa Catarina para extração dos resultados preliminares norteadores. Buscou-se comparar os grupos formados no estado catarinense com os resultados em estudos anteriores, aliado a um novo (re) agrupamento segundo a definição dos COREDEs atualmente<sup>6</sup>, portanto, distinta de quando os autores realizaram se estudo. Ainda, restringiu-se a presença de algumas variáveis mais fortemente correlacionadas entre si, encontradas nesses estudos que serviram de alicerce.

### *Fonte de dados*

O desenvolvimento desse estudo consistiu de pesquisas bibliográficas e levantamento de dados secundários, possibilitando, assim, inferir sobre os objetivos propostos. Dessa forma, para desenvolver-se, este tomou um caráter teórico-empírico. Além do pacote Microsoft Office 2003 for Windows, utilizaram-se os softwares Statistical Package for Social Sciences (SPSS) e Statistica 7.0.

Seguindo a revisão de literatura, fez-se a tabulação eletrônica dos dados coletados e a análise dos mesmos por meio da aplicação de ferramentas da análise multivariada. Tal ferramenta foi crucial, pois tornou viável o trabalho com um grande número de variáveis e observações, possibilitando a simplificação estrutural dos dados, sem comprometer a essência destes. A aplicação de técnicas multivariadas permite identificar os padrões de modernização nos municípios gaúchos e em Santa Catarina, ainda, torna possível hierarquizar-los e identificar as variáveis mais relevantes.

Foram avaliados 139 municípios do estado gaúcho sob a ótica de 14 variáveis. Os municípios fazem parte dos COREDEs essencialmente agrícolas, que conforme classificação da Fundação de Economia e Estatística (FEE) e adotado junto ao governo do estado do Rio Grande do Sul (2005) são: a) Fronteira Oeste, com 10 municípios; b) Jacuí Centro, com seis municípios; c) Missões, 21 municípios; d) Noroeste Colonial, com nove municípios; e) Celeiro<sup>7</sup>, com 20 municípios; f) Médio Alto Uruguai, com 29 municípios, g) Nordeste, com 18 municípios, h) Alto da Serra do Botucaraí, com 14 municípios; i) Alto Jacuí, com 12 municípios. Para o caso de Santa Catarina foram considerados 227 municípios, dos 260 que existiam conforme o Censo de 1995/96.

As variáveis levantadas sobre os municípios dos referidos COREDEs foram divididas em três dimensões. Uma primeira é a das variáveis agropecuárias oriundas na totalidade do Censo Agropecuário 1995/96 que medem, *stricto sensu*, os níveis de modernização da agricultura. Dentro deste grupo encontram-se variáveis que mensuram o número de estabelecimentos com assistência técnica especializada à agricultura e pecuária, o número de tratores, máquinas para colheita, utilização de adubos, irrigação, receitas agropecuárias<sup>8</sup> da agricultura e pecuária, os níveis/valores financiados, a área explorada e o efetivo bovino. Sendo objetivo analisar os possíveis e diferentes níveis de modernização nos municípios.

Uma segunda dimensão socioeconômica é medida pela população rural e pela população economicamente ativa, observações estas oriundas também do Censo 1995/96 aliadas ao índice bloco renda<sup>9</sup> dos municípios, calculados pela Fundação de Economia Estatística.

Por fim, especula-se uma questão espacial, com uma variável que mede a distância entre o município e o porto de Rio Grande, por onde escoam a maior parte da safra exportada do Rio Grande do Sul. Considerando que no estado o transporte de cargas e mercadorias está baseado praticamente do modal rodoviário, ou seja, 90% do total. As variáveis, 17 no total, estão dispostas no Quadro 1 conforme sua sigla e fonte:

Variáveis agropecuárias <sup>2</sup>	Descrição
01 ASSPEC	Assistência técnica especializada na produção animal/total de estabelecimentos no RS
02 ASSAGR	Assistência técnica especializada na produção vegetal/total de estabelecimentos no RS
03 TRATOR	Número de tratores/total de tratores no RS
04 MAQCOL	Números de máquinas para colheita/ total de máquinas para colheita no RS
05 ADUBOS	Número de estabelecimentos que utilizam adubos químicos/total de estabelecimentos no RS
06 IRRIGA	Número de estabelecimentos que utilizam sistemas de irrigação/total de estabelecimentos no RS
07 RECVEG	Valor das receitas advindas da venda de produtos vegetais/valor total no RS
08 RECANI	Valor das receitas advindas da venda de animais/valor total no RS
09 FINANC	Valor dos financiamentos nos municípios/valor total no RS
10 VALVEG <sup>1</sup>	Valor da produção vegetal/valor total no RS
11 VALANI <sup>1</sup>	Valor da produção animal de grande porte/valor total no RS
12 AREXPL	Área explorada nos municípios/total da área do RS
13 GADO	Efetivo de animais de grande porte do município/total do RS
Variáveis socioeconômicas <sup>3</sup>	Descrição
14 POP <sup>2</sup>	População rural do município/total da população rural no RS
15 PEA <sup>2</sup>	População economicamente ativa do município/total da população economicamente ativa no RS
16 RENDA <sup>1,5</sup>	Índice bloco renda no município
Variável espacial <sup>4</sup>	Descrição
17 DISTRG <sup>1</sup>	Distância do município em relação à cidade portuária de Rio Grande/total das distâncias em relação ao porto

**Quadro 1 - Variáveis utilizadas na análise de agrupamento.**

Fonte: <sup>2</sup>IBGE (1995/96), <sup>3</sup>FEE (2009), <sup>4</sup>Guia 4 rodas (2007).

Nota: <sup>1</sup>Variáveis não consideradas no agrupamento dos municípios amostrados em Santa Catarina ou descartadas.

<sup>5</sup>A variável RENDA é um índice utilizado no Índice de Desenvolvimento Socioeconômico para o Rio Grande do Sul mensurado pela FEE. O Idese é um índice sintético, inspirado no IDH, que abrange um conjunto amplo de indicadores sociais e econômicos classificados em quatro blocos temáticos: Educação; Renda; Saneamento; Domicílios; e Saúde. Todos variando entre 0 e 1.

No caso de Santa Catarina, algumas limitações levaram à exclusão da amostra municípios e variáveis. Aproximadamente, 12,70% da amostra foi retirada em função de apresentar valores nulos para variáveis como irrigação, maquinário para colheita e utilização de adubos. Ainda, para o processo de agrupamento, foi desconsiderada a variável espacial da distância, bem como o item renda. Não foi possível estimar a distância segundo os portos mais importantes de escoamento do estado, assim como não foi verificado um índice de renda como o apresentado pela FEE para o Rio Grande do Sul. De todo modo, a grande maioria dos municípios desconsiderados está no litoral catarinense, não apresentando uma economia voltada para o segmento agrícola.

*Método: análise multivariada de dados*

Kendall (1957 apud REIS, 1997) define como sendo a análise multivariada um ramo da análise estatística que verifica a relação entre grupos de variáveis dependentes e “coletadas sobre o mesmo indivíduo” (MORRISON, 1976, p. 22). As n observações sobre p variáveis

constituem a matriz de dados, em que  $n > p$ . Assim, a exigência para aplicação é que o número de variáveis seja menor que o de observações e, ainda, recomenda-se “a análise de um mínimo [...] de 17 observações” (HAIR, 2005, p. 282). O método de análise multivariada tem como objetivo otimizar a interpretação de grandes conjuntos de dados, como no caso da presente amostra extraída de dados cross-section do Censo Agropecuário de 1995/96.

A análise multivariada, de regra, na literatura está dividida em Análise de Agrupamentos (AA), Função Discriminante, Análise Fatorial, Análise das Componentes Principais, dentre outros. Estando de acordo com o sentido do método, neste trabalho optou-se prezar técnica e resultados do agrupamento dos casos propostos para o trabalho, para posterior refinamento com a utilização da Função Discriminante, dando espaço para o avanço da Análise Fatorial através da Análise das Componentes Principais.

A Análise de Agrupamentos é um método que tem como objetivo, agrupar indivíduos – casos – ou variáveis em grupos discretos. “A análise é realizada com base na similaridade ou dissimilaridade entre as variáveis a serem analisadas” (JOHNSON e WICHERN, 1992, p. 573). Sendo assim, estes serão agrupados por apresentarem características similares de acordo com algum critério de classificação.

Naturalmente, doravante o agrupamento por semelhança gera grupos com elementos homogêneos entre si, os demais grupos devem ser os mais desiguais possíveis (VALENTIN, 2000). Assim, “quanto mais próximo estão os elementos de um grupo, mais homogêneos eles serão” (HÄRDLE e SIMAR, 2003, p. 301). Contudo, é preciso que o método de aglomeração seja definido, assim como, a medida de distância entre os grupos a serem formados.

Para aplicação da AA faz-se necessária a escolha de um coeficiente que quantifique o quão parecidos dois ou mais indivíduos são. Esse coeficiente diz respeito à estimação de uma medida de similaridade/dissimilaridade entre estes indivíduos e/ou populações a serem agrupados. Na medida de similaridade, quanto maior o valor observado, mais parecidos serão os indivíduos. Enquanto na dissimilaridade, o contrário se verifica, ou seja, quanto maior o valor observado, menos parecidos estes serão.

A distância Euclidiana é a mais difundida e utilizada na análise de agrupamentos. Contudo, vale lembrar que é recomendável padronizar as variáveis antes de calcular esta distância, pois, a distância Euclidiana preserva a distância relativa quando as variáveis estão padronizadas. Para formas de agrupamento Ward e Centróide, a distância Euclidiana quadrada<sup>12</sup> é a mais recomendada (HÄRDLE e SIMAR, 2003).

Assim como na escolha entre as diversas medidas de dissimilaridade e similaridade, torna-se imprescindível a escolha do método adequado para se realizar uma AA. O “método mais difundido e comumente utilizado é a forma aglomerativa” (REIS, 1997, p. 298) e, por este motivo, será o utilizado para alcançar os objetivos propostos por este estudo. A forma aglomerativa parte de  $n$  grupos de apenas um indivíduo cada, que vão sendo agrupados sucessivamente até que se encontre um grupo que incluirá a totalidade dos  $n$  indivíduos ou variáveis em questão.

Os procedimentos hierárquicos usam cinco algoritmos aglomerativos diferentes para desenvolver os agrupamentos, o método de ligação individual, a ligação completa, a ligação média, o método de Ward e o método centróide. Para o presente estudo o algoritmo que mais se adequou aos objetivos foi o de Ward.

Assim, o método de Ward (1963) caracteriza-se pela “formação de grupos com alta homogeneidade interna e muito apropriado para variáveis quantitativas” (MINGOTI, 2005, p.178). Esse método resume-se aos seguintes passos: a) são calculadas as médias das variáveis para cada grupo; b) é calculada a distância Euclidiana quadrada; c) somam-se as distâncias para todos os indivíduos; e; d) busca-se otimizar a variância mínima dentro dos grupos.

A função objetivo que se pretende minimizar é também chamada soma dos quadrados dos erros (SQE). No princípio do processo, cada indivíduo constitui um grupo e tem  $SQE = 0$ . Logo, o passo seguinte faz agrupar os dois indivíduos que provocam um aumento mínimo no valor da soma dos quadrados dos erros, passando a existir  $n - 1$  grupos; estes  $n - 1$  grupos são então reexaminados e transformados em  $n - 2$  grupos, mas de tal modo que o aumento na função objetivo seja minimizado; o processo continua de forma sistemática até que todos os indivíduos formem um grupo apenas.

Conforme Reis (1997, p. 322), “este processo tende a formar grupos de tamanho semelhante”, contudo, como se busca grupos o mais homogêneos internamente, é preferível aceitar esta possível – e não factível – desvantagem em relação aos métodos de ligação simples e completa. Embora o método de Ward consista em agrupar indivíduos com base na soma dos quadrados entre os dois agrupamentos somados sobre todas as observações, tem-se a matriz distância Euclidiana quadrada para agrupar os indivíduos.

Estando em acordo com Mingoti (2005), em que a partição desejada é aquela que produz grupos os mais heterogêneos possíveis e de forma que os indivíduos do interior de cada grupo sejam homogêneos, o método de Ward atendeu tais prerrogativas se comparado com os demais métodos utilizados para a AA.

Quanto ao número de grupos, ainda conforme Mingoti (2005), busca-se no gráfico das distâncias de fusão entre os indivíduos o ponto em que ocorre o maior salto. Este ponto indica o número de grupos e a partição ideal ou o momento de parada do algoritmo de agrupamento.

Em outras palavras, à medida que se avança no algoritmo de agrupamento, a similaridade entre os grupos formados, vai decrescendo. E, conseqüentemente, a distância entre eles vai aumentando. Sendo assim, saltos demasiadamente grandes representam uma perda acentuada de similaridade nos grupos formados.

À medida que o nível de fusão aumenta o nível de similaridade decresce. Então, dar “o corte” no dendrograma significa traçar a Linha Fenon<sup>9</sup> que delimitará o número de grupos a se formarem. Além de buscar o menor nível de distorção que o processo aglomerativo possa trazer, é preciso ter intuição. Ainda mais, depende não só do maior salto, mas, sobretudo, do conhecimento acumulado pelo pesquisador de seu objeto de pesquisa ao decidir o ponto a ser efetuado “o corte”.

## **Análise e discussão dos resultados**

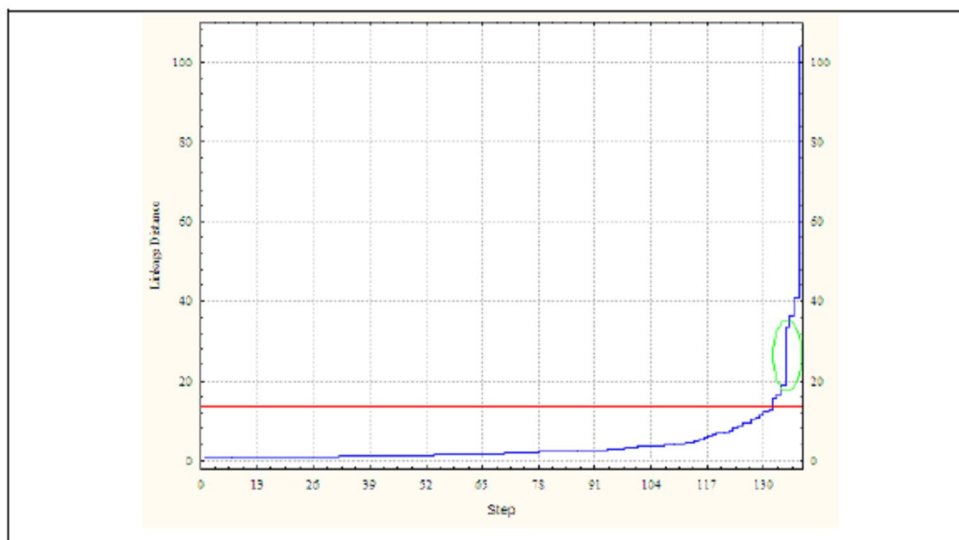
Esta seção se dedica a discutir e analisar os resultados advindos da aplicação das técnicas anteriormente expostas. Dessa forma, em um primeiro momento, agrupam-se os municípios dos COREDEs agrícolas do Rio Grande do Sul e de Santa Catarina. Esses tomaram o formato de cases que explicam a relação de proximidade dos indivíduos dentro dos grupos formados. A explicação será efetuada em um próximo estudo, com os grupos já definidos, segundo os níveis de modernização e conforme os fatores encontrados.

### *Análise de agrupamentos*

Para os COREDEs agrícolas do Rio Grande do Sul, a Análise de Agrupamentos trouxe um primeiro dendrograma, obtido no primeiro grande salto (18,82 e 33,49), no qual haveria uma considerável perda de informações, formando cerca de três grupos – elipse verde na Figura 1. O novo corte<sup>10</sup> escolhido preza pelo aumento da homogeneidade dentro dos novos grupos derivados e considera que a formação de menos grupos leva uma perda de informação pelo grande número de municípios agrupados. Um corte efetuado em uma distância inferior a



esse intervalo, em meio aos limites do início até 104,09 para a fusão total dos casos em único grupo, propicia a formação de até cinco clusters. O restrição/exclusão das variáveis referentes VALVEG e VALANI, presentes em Poerschke et al. (2008), forneceu dentro da amostra um grupo além dos quatro encontrados pelos autores. As variáveis foram restringidas em função da forte correlação entre valores e receitas com origem vegetal e animal. Pela adequação dos resultados, justifica-se o procedimento adotado, que refinou os resultados ao excluir variáveis prezando pela parcimônia. Então, têm-se cinco grupos conforme a Figura 1.



**Figura 1** - Gráfico das distâncias de fusão dos grupos nos COREDEs.

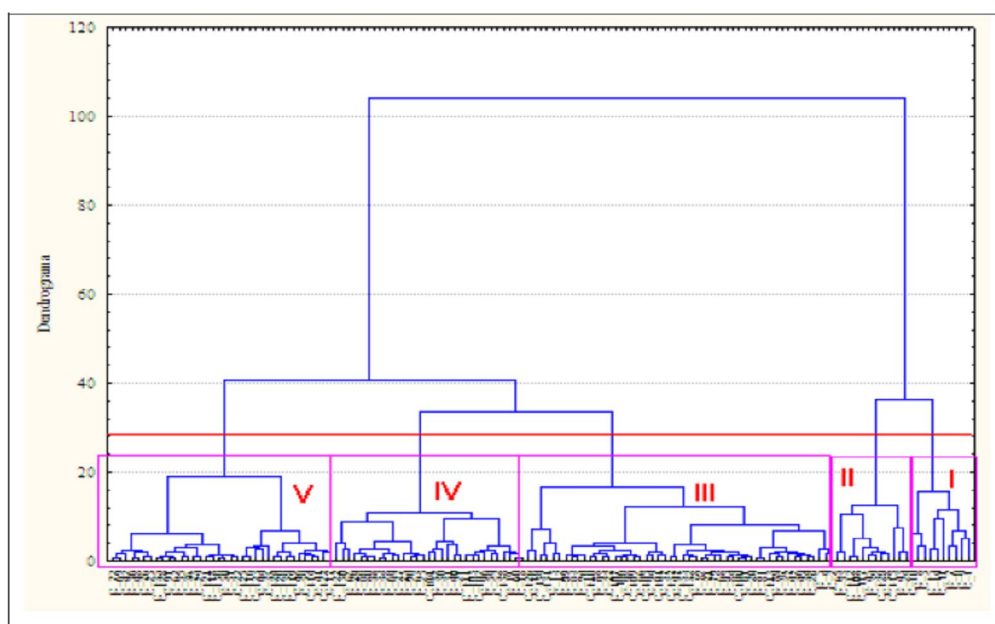
Fonte: Output fornecido pelo software Statistica 7.0.

O dendrograma da Figura 2 sugere o agrupamento ideal, bem como o corte. Assim, além do grupo I; formado por Alegrete, São Sepé, Restinga Seca, Cachoeira do Sul, Itaqui, Rosário do Sul, Santana do Livramento, São Borja, São Gabriel, Uruguaiana; tem-se o grupo II; composto por Giruá, Ajuricaba, Augusto Pestana, Ijuí, Panambi, Crissiumal, Três Passos, Sananduva, Tapejara, Cruz Alta, Ibirubá e Espumoso. O grupo III é composto por Bossoroca, Eugênio de Castro, Garruchos, Salvador das Missões, São Miguel das Missões, São Pedro do Butiá, Condor, Coronel Barros, Jóia, Pejuçara, Bom Progresso, Campo Novo, Chiapeta, Humaitá, Inhacorá, Santo Augusto, Sede Nova, Boa Vista das Missões, Dois Irmãos das Missões, Engenho Velho, Gramado dos Loureiros, Taquaruçu do Sul, Colorado, Fortaleza dos Valos, Lagoa dos Três Cantos, Não-Me-Toque, Quinze de Novembro, Saldanha Marinho, Salto do Jacuí, Santa Bárbara do Sul, Selbach, Tapera, Campos Borges, Mormaço, Nicolau Vergueiro e Vitor Graeff.

Os grupos IV e V possuem uma dimensão igualmente grande. Pois, Caibaté, Cerro Largo, Entre-Ijuís, Guarani das Missões, Porto Xavier, Roque Gonzáles, Santo Ângelo, Santo Antônio das Missões, São Luiz Gonzaga, São Paulo das Missões, Catuípe, Coronel Bicaco, Derrubadas, Redentora, São Martinho, Tenente Portela, Tiradentes do Sul, Alpestre, Erval Seco, Frederico Westphalen, Liberato Salzano, Planalto, Seberi, Água Santa, Ibiaça, Ibiraiaras, Lagoa Vermelha, São José do Ouro, Barros Cassal, Fontoura Xavier e Soledade compõe o grupo IV. O grupo V, com 50 municípios, é o maior dentre todos, pois agrega: Trindade do Sul, Vicente Dutra, Vista Alegre, André da Rocha, Barracão, Cacique Doble, Caseiros, Esmeralda, Machadinho, Maximiliano de Almeida, Paim Filho, Santo Expedito do

Sul, São João da Urtiga, Tupanci do Sul, Alto Alegre, Gramado Xavier, Ibirapuitã, Itapuca, Lagoão, São José do Herval, Barra do Guarita, Braga, Miraguaí, São Valério do Sul, Vista Gaúcha, Ametista do Sul, Caiçara, Cerro Grande, Iraí, Jaboticaba, Lajeado do Bugre, Nonoai, Novo Tiradentes, Palmitinho, Pinhal, Pinheirinho do Vale, Rio dos Índios, Rodeio Bonito, Sagrada Família, Três Palmeiras, Itacurubi, Manoel Viana, Quarai, Cerro Branco, Paraíso do Sul, Vila Nova do Sul, Dezesseis de Novembro, Pirapó, São Nicolau e Vitória das Missões.

Sendo assim, têm-se cinco grupos que compreendem na totalidade os 139 municípios propostos para o estudo. Finalmente, encerrando a etapa do agrupamento, formaram-se o grupo I com dez municípios, o grupo II com 12, o grupo III com 36 e os grupos IV e V com 31 e 50 municípios, respectivamente. Grupos estes sugeridos pelo software Statistica, seguem o mesmo padrão da Figura 2. Sendo a população do grupo I semelhante aos resultados de Poerschke et al. (2008). Esse grupo apresentou os principais municípios gaúchos produtores de arroz.

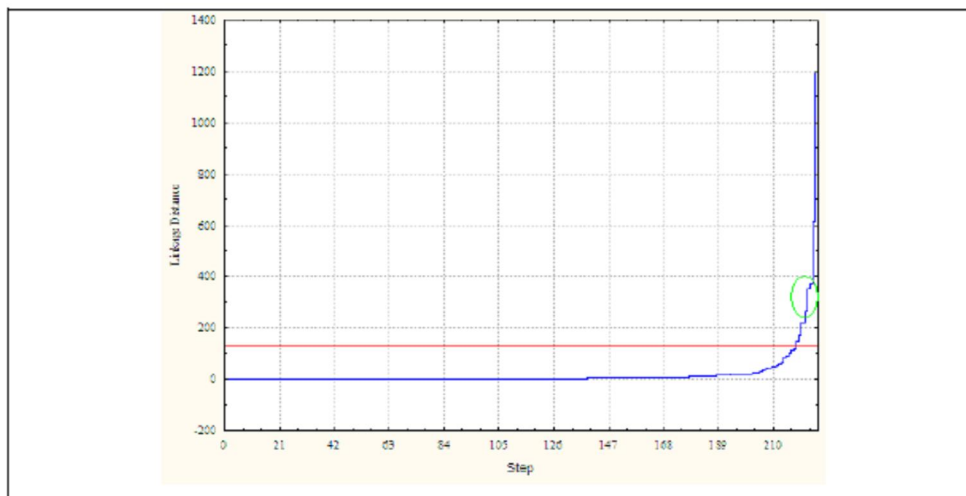


**Figura 2** – Dendrograma dos COREDEs.

Fonte: Output fornecido pelo software Statistica 7.0.

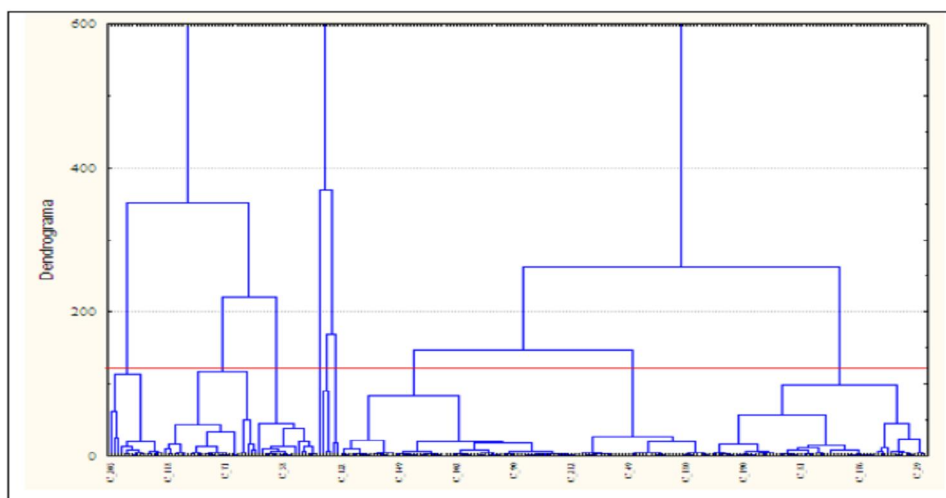
Para Santa Catarina, os mesmos critérios foram adotados. Ainda, a rigidez desses não foi suficiente para fornecer grupos consistentes na amostra dos 227 municípios considerados. A Figura 3 salienta o salto escolhido e, embora inferior ao grande salto, não se mostra consistente. Houve municípios que se apresentaram ímpares, como Nova Itaberaba Arabutã, Caçador, Mafra, Abelardo Luz e Curitibaanos. Ainda, formou-se um total de oito grupos desequilibrados<sup>15</sup> entre si. Esses resultados prévios sugerem a necessidade do refinamento da amostra em função das Secretarias de Desenvolvimento Regional (SDRs) com predominância da atividade agrícola. Nesse sentido, a não existência de um trabalho semelhante ao Rumos 2015 (2006) limita a possibilidade de prosseguir na análise e cruzar com dados mais consistentes já apresentados pela amostra dos COREDEs gaúchos. Essa prévia classificação para seleção da amostra catarinense tende a trazer resultados mais conclusivos e passíveis de comparação. Ainda que, municípios como Turvo, Massaranduba, Guaramirim, Morro Grande, Jacinto Machado, Meleiro, Jaguaruna, dentre outros, formaram um subgrupo dentro de um

dos grandes grupos entre os oito encontrados. Essa convergência do resultado nesse ponto sugere a possibilidade e plausibilidade de se aceitar que o mesmo processo possa ser verificado em Santa Catarina, embora o método de plantio seja diferenciado, pois esses municípios são os principais produtores de arroz no estado catarinense. Assim, é necessário seleccionar uma nova amostra segundo aquelas Secretarias essencialmente agrícolas.



**Figura 3** - Gráfico das distâncias de fusão dos grupos em Santa Catarina.  
 Fonte: Output fornecido pelo software Statistica 7.0.

O subgrupo mencionado assume posição de destaque dentro dos produtores de arroz do estado catarinense. Esses municípios, assim como no Rio Grande do Sul, se mantiveram entre os principais produtores ao longo da última década, conforme os dados da Tabela 2. O dendrograma pode ser vislumbrado na Figura 4, o qual salienta o elevado número de grupos formados, bem como a disparidade dimensional dos mesmos.



**Figura 4** – Dendrograma de Santa Catarina.  
 Fonte: Output fornecido pelo software Statistica 7.0.

O elevado número de municípios da amostra impossibilita expor um maior detalhamento visual.

**Tabela 2** - Produção de arroz em casca – municípios selecionados de Santa Catarina - comparativo 1995 e 2006 (toneladas)

Município	Arroz (base casca)				
	1995	(%)		2006	(%)
Turvo	57.750	43,81	Forquilhinha	73.8750	7,33
Massaranduba	55.240	41,90	Turvo	69.2380	7,23
Meleiro	48.000	36,41	Meleiro	65.8000	6,06
Guaramirim	33.786	25,63	Nova Veneza	53.9000	5,43
Nova Veneza	33.610	25,49	Guaramirim	52.8000	4,91
Forquilhinha	30.259	22,95	Massaranduba	47.2420	4,74
Jaguaruna	26.750	20,29	Jacinto Machado	43.0950	4,50
Jacinto Machado	24.000	18,20	Tubarão	34.5800	3,33
Morro Grande	24.000	18,20	Gaspar	33.3200	3,29
Tubarão	23.100	17,52	Jaguaruna	30.1620	3,20
Santa Catarina	710.422		Santa Catarina	1.071.559	
Brasil	11.226.064		Brasil	11.526.685	

Fonte: IBGE (2009).

### Considerações finais

Tido como objetivo geral agrupar municípios o mais homogêneos possível, a Análise de Agrupamentos mostrou seu importante papel para a pesquisa econômica. Mensurar o grau de modernização dos municípios, em vista de seu elevado número, foi possível pela utilização das técnicas de análise multivariada de dados socioeconômicos, revelando, assim, uma excelente ferramenta para pesquisas.

Nos resultados obtidos, a realidade da modernização agrícola brasileira – excludente e conservadora – se repetiu no estado do Rio Grande do Sul explicitamente, pois, uma minoria de municípios, dez, dispõe de elevado grau de modernização na agricultura. Essa realidade demonstra a concentração da agricultura, dita moderna, no grupo I. Por outro lado, revelou que 129 municípios da amostra não têm uma agricultura considerada moderna, além de não ser intensivos na produção de arroz, que demanda muita irrigação e inversão de capital.

O mesmo comportamento e adequação da amostragem feita para Santa Catarina não foi consistente com os resultados da metodologia, a qual ficou limitada pelo menor número de variáveis disponíveis (11), bem como o elevado número de municípios (227) e o grau de disparidade entre os mesmos conforme revelou o dendrograma. Formaram-se muitos grupos, oito no total, inclusive grupos com um único elemento. Sendo o forte do método de Ward evitar esse tipo de forma de agrupamento, os resultados impossibilitaram o presente estudo de cruzar os resultados entre os estados sob a ótica da produção de arroz. De todo modo, os resultados revelaram o próximo passo a seguir para atingir os objetivos, ou seja, a necessidade de estimar as SDRs que abrangem municípios relacionados, predominantemente, com a atividade agropecuária, indústria ou um mix dos dois no estado catarinense.

Finalmente, os objetivos foram em parte atingidos com o refinamento dos resultados já conhecidos, agora com a nova formatação dos COREDES e com a eliminação das variáveis que causaram problemas de multicolinearidade. Os grupos formados foram consistentes e ressaltaram as características objetivadas quanto a especialização produtiva de arroz. Para Santa Catarina, o estudo padeceu da ausência de variáveis importantes, bem como, de estudos preliminares de classificação e agrupamento dos municípios do estado. Como sugestão futura,

podem-se cruzar os dados do novo Censo Agrícola 2007/08, para ver se permanece ou não o padrão de concentração em determinado grupo.

## Notas

<sup>1</sup> Ver Freitas et al. (2007) e Poerschke et al. (2008).

<sup>2</sup> Referências ao ano safra, ou seja, considera-se o período entre março de 2008 até fevereiro de 2009.

<sup>3</sup> Para discussão da metodologia ver Rummel (1967).

<sup>4</sup> COREDE que é formado por Alegrete, Itacurubi, Itaqui, Manoel Viana, Quaraí, Rosário do Sul, Santana do Livramento, São Borja, São Gabriel e Uruguaiana.

<sup>5</sup> Cachoeira do Sul, Cerro Branco, Paraíso do Sul, Restinga Seca, São Sepé e Vila Nova do Sul.

<sup>6</sup> Decreto nº 45.436, de 09 de janeiro de 2008, acrescenta aos 26 Conselhos Regionais de Desenvolvimento anteriores mais dois COREDEs: o Vale do Jaguarí e o Celeiro.

<sup>7</sup> Criado em 10 de janeiro de 2008.

<sup>8</sup> Os valores não foram deflacionados, pois, não se tratando de um comparativo em dois momentos, ou mesmo, de séries temporais, os resultados não se alteraram.

<sup>9</sup> Este indicador medido pela FEE leva em consideração o Valor Adicionado Bruto (VAB) no comércio, alojamento e alimentação e geração de renda per capita. O presente trabalho versa sobre os COREDEs essencialmente agrícolas, dessa forma justificando a utilização do bloco renda dos COREDEs predominantemente agrícolas.

<sup>10</sup> Definida como  $D_{k,i}^2 = \sum_{r=1}^k (X_{r,i} - X_{r,D})^2$ .

<sup>11</sup> Linha Fenon, ou linha de corte, marca a distância onde ocorre o maior salto no gráfico das distâncias de fusão, traçada em vermelho nas figuras que seguem.

<sup>12</sup> Linha Fenon em vermelho da Figura 01 indica o instante do corte no processo de agrupamento, entre 16,62 e 18,82.

## Referências

CAPORAL, Francisco Roberto; COSTABEBER, Jose Antônio. Agroecologia e Extensão Rural: contribuições para a promoção do desenvolvimento rural sustentável. Brasília: MDA/SAF/DATER-ICCA, 2004.

COMPANHIA NACIONAL DE ABASTECIMENTO (Conab). Acompanhamento da safra brasileira: grãos. Brasília: nov. 2009.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (Embrapa) / Centro Nacional de Pesquisa de Arroz e Feijão (CNPAF). Banco de dados.

FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA (FEE). Banco de Dados: IDESE. Disponível em: <<http://www.fee.tche.br>>. Acesso em: nov. 2009.

FREITAS, Clailton. A.; PAZ, Marlon V.; NICOLE, Danieli S. Analisando a modernização da agropecuária gaúcha: uma aplicação da análise fatorial e cluster. Análise Econômica, Porto Alegre, v. 25, p. 121-149, 2007.

GUIA 4 RODAS. Disponível em: <<http://www.guia4rodas.abril.com.br/mapasbrasi>>. Acesso em: 10 abr. 2007.

HAIR, J. F. Análise multivariada de dados. 5ª ed. Porto Alegre: Bookman, 2005.

HÄRDLE, W.; SIMAR, L. Applied Multivariate Statistical Analysis. Berlin and Louvain-la-Neuve, 2003.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 1 abr. 2009.

\_\_\_\_\_. Censo Agropecuário 1995-1996 – Rio Grande do Sul e Santa Catarina. v. 20, 21 e 22. 1998.

JOHNSON, R. A.; WICHERN, D. W. Applied Multivariate Statistical Analysis. 3ª ed. New Jersey: Prentice Hall, 1992.

KENDALL, M. G. A Course in Multivariate Analysis. London: Griffin, 1957.

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

MORRISON, D. F. Multivariate statistical methods. New York: Mc Graw Hill, 1976.

POERSCHKE, Rafael P.; MOREIRA JUNIOR, Fernando J.; FREITAS, Clailton A. Um retrato da modernização agropecuária nos COREDES agrícolas do Rio Grande do Sul: Uma aplicação da análise multivariada. Anais do 4º Encontro de Economia Gaúcha, 2008, Porto Alegre.

POERSCHKE, Rafael P. Criação e Desvio de Comércio no Mercosul: O caso do arroz em casca. (Dissertação de Mestrado) - Universidade Federal de Santa Maria, 2008.

REIS, E. Estatística Multivariada. Lisboa: Edições Silabo, 1997.

RUMMEL, R. J. Understanding factor analysis. Journal of Conflict Resolution, v. 11, n. 4, p. 444-480, 1967.

SECRETARIA DE COORDENAÇÃO E PLANEJAMENTO DO RS. RUMOS 2015: estudo sobre desenvolvimento regional e logística de transportes no Rio Grande do Sul. Porto Alegre, 2006.

SILVA, José Graziano da. Tecnologia e Agricultura Familiar. Campinas: UNICAMP, 1999.

VALENTIN, J. Ecologia numérica - uma introdução à análise multivariada de dados ecológicos. Rio de Janeiro: Interciência, 2000. 117 p.

VEIGA, Jose Eli da. O desenvolvimento agrícola: uma visão histórica. São Paulo: Editora da Universidade de São Paulo, 1991.

WARD, J. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 1963, pp. 236-244.