

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
PROGRAMA DE PÓS GRADUAÇÃO EM ESTATÍSTICA E
MODELAGEM QUANTITATIVA



Mariangela Amaral e Silva

**INDICE DE DESENVOLVIMENTO RURAL PARA O COREDE
CENTRAL DO RIO GRANDE DO SUL (2010): ANÁLISE
MULTIVARIADA**

Santa Maria, RS

2020

Mariangela Amaral e Silva

**INDICE DE DESENVOLVIMENTO RURAL PARA O COREDE CENTRAL DO
RIO GRANDE DO SUL (2010): ANÁLISE MULTIVARIADA**

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa.**

Orientador: Prof. Dr. Adriano Mendonça Souza

Santa Maria, RS

2020

Mariângela Amaral e Silva

**ÍNDICE DE DESENVOLVIMENTO RURAL PARA O COREDE CENTRAL DO
RIO GRANDE DO SUL (2010): ANÁLISE MULTIVARIADA**

Monografia apresentada ao Curso de Especialização em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Especialista em Estatística e Modelagem Quantitativa**.

Aprovado em 26 de março de 2020:



Adriano Mendonça Souza, Dr. UFSM
(Presidente/Orientador)



Roselaine Ruviano Zanini, Dr^a. (UFSM)



Luciana Flores Jacobi, Dr^a. (UFSM)

Santa Maria, RS
2020.

DEDICATÓRIA

Dedico aos meus filhos Cauê, Ian e Enzo,
que são a obra prima de Deus na minha vida.

AGRADECIMENTOS

Agradeço primeiramente a Deus e aos meus pais pela possibilidade de estar viva e ter capacidade de realizar esta pesquisa. Obrigada Deus, obrigada mamãe e obrigada papai!

Agradeço aos professores, colegas e à coordenação, na pessoa da Prof. Luciane Flores Jacobi, que não mediu esforços em dar seu apoio todas as vezes que a procurei para organizar os horários para que eu pudesse cursar, devido ao trabalho e em particular à dedicada orientação do Prof. Adriano Mendonça Souza, que com seu bom humor sempre me recebeu e me orientou com extrema competência.

Quero dedicar este estudo às pessoas que são as principais responsáveis por eu estar vencendo mais esta etapa, que são meus queridos pais, e meus filhos Cauê, Ian e Enzo, aos quais agradeço imensamente pela compreensão às várias ausências ocorridas durante esta empreitada e também pelo eterno estímulo recebido de todos os familiares que me permitiram chegar ao final com sucesso.

Agradeço muito a colega Helen pelas várias horas de estudo, pelo constante apoio e amizade, e aos demais colegas os quais fizeram deste curso uma grande família. Agradeço a todas as pessoas, coisas e fatos do universo, que de uma forma ou de outra, colaboraram para a conclusão deste trabalho.

RESUMO

ÍNDICE DE DESENVOLVIMENTO RURAL PARA O COREDE CENTRAL DO RIO GRANDE DO SUL (2010): ANÁLISE MULTIVARIADA

AUTORA: Mariangela Amaral e Silva
ORIENTADOR: Prof. Dr. Adriano Mendonça Souza

O desenvolvimento rural é importante por tratar-se de um processo necessário para promover as transformações sociais, econômicas e ambientais no espaço rural, a fim de melhorar o bem-estar social e econômico e a qualidade de vida das populações do campo. Nesta monografia, o principal objetivo foi construir um Índice de Desenvolvimento Rural (IDR) para o COREDE Central do Rio Grande do Sul, composto por 19 municípios. Apresentou-se os municípios mais homogêneos e quais fatores fizeram parte da composição do desenvolvimento rural e as suas diferenças. A metodologia utilizada na pesquisa foi Análise de Cluster, Análise Fatorial e Análise dos Componentes Principais, onde foi feita a análise de agrupamento para verificar a similaridade das variáveis e se chegou a dois grupos reduzido a nove variáveis, a partir das quais se utilizou para a análise fatorial e análise dos componentes principais. Estas resultaram em dois autovalores maiores que 1, que explicam 87,30% da variância total dos dados e chegou-se a dois IDRs. Os resultados retratam o perfil da região com grande dependência do setor agropecuário, com baixa diversidade da base econômica e certa concentração fundiária. Os dois IDRs mostram essas características regionais, quando em meio a múltiplas variáveis pertencentes a quatro dimensões (social, demográfica, econômica e ambiental), de desenvolvimento, chega-se ao destaque para a produção de soja (SOJ) e arroz (ARR) e elevado valor adicionado bruto (VAB) no setor agropecuário, predominando a dimensão econômica. Sendo que os demais setores, comércio, serviços e indústrias são fomentados pelo setor rural. Os municípios de Júlio de Castilhos e Tupanciretã são representados pelo IDR2 construído com as variáveis: valor da produção de soja e alto VAB, demonstrando o predomínio da dimensão econômica e confirmando assim a realidade onde esses dois municípios figuram como os maiores produtores de soja do estado do Rio Grande do Sul, e os municípios de Agudo e Santa Maria são representados pelo IDR1 formado pelas variáveis: número de estabelecimentos rurais com controle de pragas e/ou doenças (ECP) e população rural ocupada no ambiente rural (POC) demonstrando o predomínio da dimensão ambiental e demográfica.

Palavras-chave: Desenvolvimento Rural. Análise Multivariada. Índice.

ABSTRACT

RURAL DEVELOPMENT INDEX FOR THE RIO GRANDE DO SUL CENTRAL COREDE (2010): MULTIVARIATE ANALYSIS

AUTHOR: Mariangela Amaral E Silva
ADVISOR: Prof Dr. Adriano Mendonça Souza

Rural development is important because it is a necessary process to promote social, economic and environmental changes in rural areas, in order to improve the social and economic well-being and the quality of life of rural populations. In this monograph, the main objective was to build a Rural Development Index (IDR) for the COREDE Central of Rio Grande do Sul, composed of 19 municipalities. The most homogeneous municipalities were presented and what factors were part of the composition of rural development and their differences. The methodology used in the research was Cluster Analysis, Factor Analysis and Principal Component Analysis, where the cluster analysis was performed to verify the similarity of the variables and two groups were reduced to nine variables, from which it was used for the analysis. factor analysis and analysis of the main components. These resulted in two eigenvalues greater than 1, which explain 87.30% of the total variance of the data and two IDRs were arrived at. The results show the profile of the region with great dependence on the agricultural sector, with low diversity of the economic base and a certain concentration of land ownership. The two IDRs show these regional characteristics, when in the midst of multiple variables belonging to four dimensions (social, demographic, economic and environmental), of development, the highlight is the production of soy (SOJ) and rice (ARR) and high gross value added (GVA) in the agricultural sector, with the economic dimension predominating. The other sectors, commerce, services and industries are encouraged by the rural sector. The municipalities of Júlio de Castilhos and Tupanciretã are represented by the IDR2 constructed with the variables: value of soybean production and high GVA, demonstrating the predominance of the economic dimension and thus confirming the reality where these two municipalities figure as the largest soybean producers in the state of Rio Grande do Sul, and the municipalities of Agudo and Santa Maria are represented by the IDR1 formed by the variables: number of rural establishments with pest and / or disease control (ECP) and rural population employed in the rural environment (POC) demonstrating the predominance the environmental and demographic dimension.

Keywords: Rural Development. Multivariate analysis. Index.

LISTA DE TABELAS

Tabela 1- Autovalores da matriz e variância explicada das correlações para as variáveis de desenvolvimento rural dos municípios do COREDE central/RS – ano 2010.....	39
Tabela 2 -Cargas fatoriais das variáveis de desenvolvimento rural dos municípios do COREDE central/RS após rotação ortogonal e comunalidades – ano 2010	40
Tabela 3 -Contribuição das variáveis baseada nas correlações	43

LISTA DE QUADROS

Quadro 1- Procedimentos para análise de agrupamento (AA)	20
Quadro 2 - Classificação e enquadramento metodológico	30
Quadro 3 - Variáveis de desenvolvimento rural utilizadas – ano 2010.....	31

LISTA DE FIGURAS

Figura 1- Mapa do COREDE Central (2010)	18
Figura 2 - Passos para realizar a análise fatorial	23
Figura 3 - Esquema para análise de componentes principais	25
Figura 4 - As 15 variáveis relacionadas aos 19 municípios do COREDE Central/RS	35
Figura 5 - Dendrograma com 10 variáveis, após a retirada de variáveis com mesmo grau de similaridade	36
Figura 6 - Dendrograma com 9 variáveis, após a retirada de variáveis com o mesmo grau de similaridade	37
Figura 7 - Dendrograma com os 19 municípios do COREDE Central	38
Figura 8 - Gráfico representa a relação entre fatores (Fator1 x Fator2) .	41
Figura 9 - Gráfico de distribuição da nuvem de pontos (os municípios) ..	44

SUMÁRIO

1. INTRODUÇÃO	11
1.1 OBJETIVOS	13
1.1.1 Objetivo geral	13
1.1.2 Objetivos específicos:	13
1.2. JUSTIFICATIVA	13
1.3 ESTRUTURA DO TRABALHO	14
2. REFERENCIAL TEÓRICO	15
2.1. DESENVOLVIMENTO RURAL	15
2.2. CONSELHO REGIONAL DE DESENVOLVIMENTO - COREDE	17
2.3 ANÁLISE MULTIVARIADA	19
2.3.1 Análise de Agrupamento (Cluster)	20
2.3.2 Análise Fatorial	22
2.3.3 Análise de Componentes Principais – ACP	25
3. MATERIAIS E MÉTODOS	29
3.1 CARACTERIZAÇÃO DA PESQUISA.....	29
3.2 ÁREA DE ESTUDOS E VARIÁVEIS ANALISADAS	30
3.3 MÉTODO DE ANÁLISE	31
4. ANÁLISE DOS RESULTADOS	34
4.2 ANÁLISE DE CLUSTER	34
4.3 ANÁLISE FATORIAL	39
4.4 A CONSTRUÇÃO DO IDR VIA ANÁLISE DE COMPONENTES PRINCIPAIS.....	42
4.4.1 Índices de desenvolvimento rural	43
5. CONSIDERAÇÕES FINAIS	46
REFERÊNCIAS	48

1. INTRODUÇÃO

A diversidade e a heterogeneidade do espaço rural são conhecidas como dimensões do desenvolvimento rural. (CONTERATO E SCHNEIDER, 2010). Kageyama (2008) afirma que é necessário avançar na diferenciação entre os indicadores e dimensões dos fatores e das causas da desigualdade no desenvolvimento rural para o que são as suas consequências.

As mudanças que se processaram na agricultura brasileira, principalmente a partir do final da década de 1960, caracterizaram uma redefinição das relações entre a agricultura e a indústria, dando origem a um novo padrão de produção agrícola, o uso de sementes selecionadas e de insumos químicos, irrigação, mecanização agrícola e variedades genéticas das culturas que mais se adaptaram ao ambiente das regiões brasileiras. Entretanto, os impactos não foram uniformes em todas as regiões do país e nem em todos os estratos de produtores: os incentivos privilegiaram o grande capital agrícola, enquanto estimulavam a expropriação e a expulsão do homem do campo; intensificou-se a concentração fundiária; foram privilegiadas as regiões mais desenvolvidas e os grandes produtores rurais, argumentam (MELO E PARRÉ, 2007).

Sobre a ideia de desenvolvimento coexistem várias visões distintas que disputam entre si os seus significados. De acordo com Kageyama (2004), “o desenvolvimento – econômico, social, cultural, político – é um conceito complexo e só pode ser definido por meio de simplificações, que incluem *decomposição* de alguns de seus aspectos e *aproximação* por algumas formas de medidas.”

Segundo o relatório de desenvolvimento humano (PNUD¹, 2015), desde 1990 o relatório afirma que o desenvolvimento tem como propósito básico ampliar as escolhas das pessoas, proporcionando um ambiente onde elas possam gozar de uma vida longa, saudável e criativa. Para isso, o avanço de uma população deve ser medido por suas características sociais, culturais e políticas, e não apenas pela sua dimensão econômica.

Dentro desse contexto, a partir da criação do Índice de Desenvolvimento Humano (IDH) pelos economistas Mahbub ul Haq, do Paquistão, e Amartya Sen, indiano ganhador do Prêmio Nobel de Economia de 1998, deixou-se de

¹ Programa das Nações Unidas para o Desenvolvimento

considerar apenas o PIB *per capita* como medida de desenvolvimento, pois o IDH, além de considerar o aumento da renda, leva em conta outros valores que também fazem parte da expansão das opções humanas de um indivíduo. Com a criação desse índice, os autores vem ao encontro da visão de que a explicação de desenvolvimento não deve ser reduzida apenas a sua dinâmica econômica, que seria diluída no crescimento econômico, mas considerar outras variáveis sociais, culturais, políticas e ambientais, que vão além de um simples aumento da riqueza avançando para a construção de um bem-estar econômico, social e ambiental das populações.

Nessa perspectiva, Conterato, Schneider, Waquil (2009) definem o desenvolvimento rural como sendo um processo que é resultado de ações articuladas com o objetivo de promover mudanças sociais, econômicas e ambientais no espaço rural, tendo a intenção de melhorar a renda, o bem-estar e a qualidade de vida das populações rurais.

O principal objetivo deste trabalho foi construir um Índice de Desenvolvimento Rural (IDR) para o COREDE Central do Rio Grande do Sul, composto por 19 municípios. Além disso, apresentou-se os municípios mais homogêneos e quais fatores compõem o desenvolvimento rural e as suas diferenças. A construção do IDR teve a intenção de captar de forma simplificada (o índice) as diferenças regionais do desenvolvimento rural e corroborar teoricamente com a compreensão empírica, pois de acordo com Conterato, Schneider, Waquil (2009) “o desenvolvimento, enquanto conceito e processo complexo, inevitavelmente traz consigo inúmeras formas de apreensão e determinantes das suas manifestações.”

Para alcançar o almejado, foi aplicada a Análise Multivariada de dados, onde, primeiramente, foi feita a Análise de Agrupamento para verificar as variáveis semelhantes, após uma Análise Fatorial para saber quais fatores compõem o desenvolvimento rural e, por último, a construção do IDR por meio dos Componentes Principais.

De acordo com a abordagem sobre a perspectiva de promover mudanças sociais, econômicas e ambientais no ambiente rural buscando a melhora na vida das populações do campo, essa pesquisa construiu o Índice de Desenvolvimento Rural (IDR) para o COREDE Central do Rio Grande Sul.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Construir o índice de desenvolvimento rural (IDR) para o COREDE Central/RS.

1.1.2 Objetivos específicos:

- a) Verificar quais as variáveis e os municípios homogêneos aplicando a análise de agrupamento;
- b) Identificar fatores comuns aplicando Análise Fatorial;
- c) Encontrar o IDR por meio da aplicação de Análise dos Componentes Principais.

1.2. JUSTIFICATIVA

A desigualdade de renda é um problema relevante para um país e no Brasil, o índice de desigualdade é elevado, o que mostra que existe grande concentração da renda e que quanto maior for essa concentração, maior é a privação do povo em saúde, educação, expectativa de vida, de acordo com o Programa das Nações Unidas para o Desenvolvimento (PNUD, 2016).

Segundo Kageyama (2004), já foi publicado pelo PNUD, desde 1990 que fazer análise considerando apenas do Produto Interno Bruto *per capita* é insuficiente como medida de desenvolvimento. É necessário incorporar na análise outras variáveis, tendo em vista que a análise de desenvolvimento inclui o aumento de bem-estar da sociedade, que vai além da simplificação do aumento da renda.

Esta pesquisa vem com a intenção de continuidade ao estudo das desigualdades com uma ampliação para a construção do Índice de Desenvolvimento Rural para o COREDE Central-RS, região onde está instalado o IFFar-JC – Instituto Federal Farroupilha – campus Júlio de Castilhos. A temática justifica-se por tratar-se de uma Instituição de ensino com cursos do setor agropecuário em uma região de perfil rural. Sendo assim, percebe-se a importância de um estudo nesse setor.

Quanto à relevância para a sociedade, este estudo pode servir para dar suporte na região, pois o IDR vai dar um panorama com relação ao desenvolvimento. A partir do estudo das variáveis para a construção do índice, à luz do processo histórico de desenvolvimento rural dessa região, foi possível buscar alternativas que incorporem a região no processo de desenvolvimento, dando suporte para implantação de políticas, planos, programas e ações que estruturam um modelo de desenvolvimento rural sustentável.

1.3 ESTRUTURA DO TRABALHO

O presente trabalho está estruturado em cinco capítulos. O primeiro aborda o tema principal, o problema da pesquisa, o objetivo geral e os específicos, a justificativa e a estrutura do trabalho. O segundo capítulo apresenta a conceituação e aprofundamento do desenvolvimento rural que é feito através da revisão de literatura de trabalhos científicos de autores renomados no tema. No terceiro capítulo, faz-se apresentação da metodologia utilizada para construir o IDR e demais análises. No quarto capítulo são expostos os resultados e, por fim, no capítulo 5 as considerações finais.

2. REFERENCIAL TEÓRICO

Este tópico dá suporte teórico à pesquisa e encontra-se dividido em duas etapas. Na primeira, explicitam-se da conceituação e ambientação referente ao processo de desenvolvimento rural. Num segundo momento é apresentado sobre COREDE sendo detalhado o COREDE Central que é a área de estudo deste trabalho. É importante salientar que a pesquisa foi baseada nos estudos de diversos autores, porém, merecem destaque (KAGEYAMA, 2004; CONTERATO; SCHNEIDER; WAQUIL, 2009; MELO E PARRÉ, 2007; HAIR et al, 2009, VICINI et al, 2018), entre outros.

2.1. DESENVOLVIMENTO RURAL

O exame da literatura disponível sobre o tema revelou alguns estudos que conceituaram e avaliaram o desenvolvimento rural.

A Ciência Econômica trata o desenvolvimento econômico como sendo o crescimento econômico, mensurado pelo produto nacional bruto per capita (PIB), acompanhado pela melhoria do padrão de vida da população e por alterações fundamentais na estrutura de sua economia. Assim pode-se entender que o desenvolvimento de cada país ou região depende de suas características próprias (situação geográfica, passado histórico, extensão territorial, população, cultura e recursos naturais). De maneira geral, as mudanças que caracterizam o desenvolvimento econômico consistem no aumento da atividade industrial em comparação com a atividade agrícola, migração da mão-de-obra do campo para as cidades, redução das importações, aumento das exportações, menor dependência de auxílio externo.

Na análise de Vieira e Maimon (1993), o conceito de progresso está na base da ótica desenvolvimentista que perpassa de forma temporal por quatro fases: desenvolvimento enquanto sinônimo de crescimento; desenvolvimento enquanto etapa; desenvolvimento enquanto processo de mudança estrutural e desenvolvimento sustentável.

Os dois primeiros enfoques prevalecem nos anos 50 e 60, quando a sociedade é constituída de unidades econômicas que seguem processos mecanicistas e cujas leis são conhecidas cientificamente. Os objetivos do desenvolvimento são a maximização do produto nacional bruto e da renda per

capita, obtidos através da maximização da eficiência econômica. O terceiro enfoque ganhou repercussão com os trabalhos de centro-periferia da CEPAL², a partir do final da década de 60. E na década de 80, a ênfase passa a ser do Desenvolvimento Sustentado, com base no crescimento econômico, equidade social e harmonia ambiental.

O Programa de Desenvolvimento Rural da OCDE³, lançado em 1991 estabeleceu três etapas para a geração de indicadores de desenvolvimento rural e classificou as regiões estudadas em três tipos: predominantemente rurais, predominantemente urbanizadas e significativamente rurais. O indicador básico que define as unidades territoriais rurais e urbanas e permite classificar as regiões é a densidade populacional. Após, estabeleceu um conjunto de indicadores demográficos, econômicos, sociais e ambientais, e a partir dos dados coletados foram produzidas as séries estatísticas para comparar e analisar as tendências do desenvolvimento rural. (OECD, 1996)

Kageyama (2004) partiu do conjunto de indicadores que a OCDE (1991) utilizou para descrever o desenvolvimento rural que, segundo a autora, trata-se de um conceito complexo e multissetorial, para construir e classificar o índice de desenvolvimento rural (IDR) para os municípios de São Paulo, procurando reunir numa medida síntese aspectos populacionais, econômicos, sociais e ambientais que permitem classificar os municípios de São Paulo segundo seu grau de desenvolvimento rural.

Para construir o IDR, Kageyama (2004), calculou quatro índices parciais representando quatro aspectos (População, economia, bem-estar social e meio ambiente), cada um dos índices é composto pela média de indicadores simples, e o IDR, é a média aritmética dos quatro índices parciais.

Waquil e Schneider (2009) analisaram as desigualdades regionais de desenvolvimento rural construindo o IDR para três microrregiões do Rio Grande do Sul (Caxias do Sul, Frederico Westphalen e Cerro Largo). Consideraram para a composição do IDR as dimensões social, econômica, demográfica, político-institucional e ambiental.

² A Comissão Econômica para a América Latina e o Caribe criada em 1948 pelo Conselho Econômico e Social das Nações Unidas com o objetivo de incentivar a cooperação econômica entre os seus membros.

³ Organisation de Coopération et Développement Économiques.

Concluíram que o índice proposto demonstrou que o desenvolvimento rural é um processo multifacetado e multidimensional. Waquil e Schneider (2009) afirmam que as desigualdades regionais de desenvolvimento rural caracterizadas em suas pesquisas permitiram concluir com algum grau de confiança que os processo de mudanças social no meio rural não ocorrem com a mesma intensidade e ao mesmo tempo em todos os lugares. Com relação as unidades territoriais investigadas, pode-se considerar que o desenvolvimento rural é intra-regionalmente harmônico em suas dimensões e inter-regionalmente desigual.

Nessa pesquisa será feita a análise do desenvolvimento rural da região do COREDE Central/RS, sendo assim torna-se necessário o conhecimento sobre essa temática, é o que será tratado no próximo tópico.

2.2. CONSELHO REGIONAL DE DESENVOLVIMENTO - COREDE

Os Conselhos Regionais de Desenvolvimento - COREDE, criados oficialmente pela Lei 10.283 de 17 de outubro de 1994, tendo como função formar um fórum de discussão com a finalidade de construir políticas e ações a fim de viabilizar o desenvolvimento regional.

De acordo com o Atlas Socioeconômico do RS (2020) seus principais objetivos são a promoção do desenvolvimento regional harmônico e sustentável; a melhoria da eficiência na aplicação dos recursos públicos e nas ações dos governos para a melhoria da qualidade de vida da população e a distribuição equitativa da riqueza produzida; o estímulo a permanência do homem na sua região e a preservação e recuperação do meio ambiente.

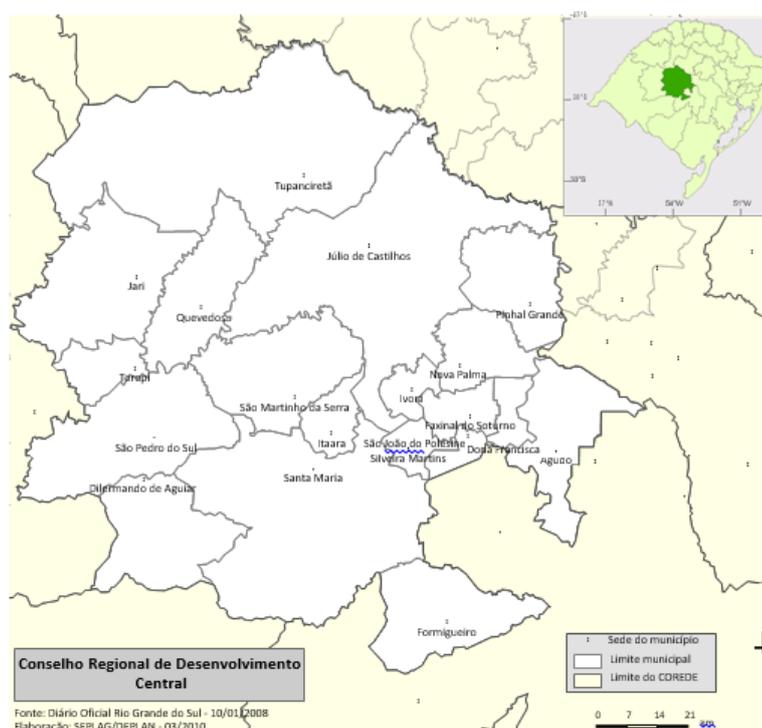
A divisão regional, inicialmente composta por 21 regiões, foi alterada em 1998 com a criação do 22º COREDE – Metropolitano Delta do Jacuí e, em 2003, com a criação dos COREDEs Alto da Serra do Botucaraí e Jacuí Centro. Em 2006 foram criados os COREDEs Campos de Cima da Serra e Rio da Várzea. E, finalmente, em 2008, através do Decreto 45.436, foram criados os COREDEs Vale do Jaguari e Celeiro. O Estado conta, atualmente, com 28 Conselhos Regionais de Desenvolvimento. Este estudo pretende considerar na análise para a construção do IDR o COREDE Central.

O COREDE Central está composto por 19 municípios pertencentes a microrregião Central do Rio Grande do Sul: Agudo, Dilermando de Aguiar, Dona

Francisca, Faxinal do Soturno, Formigueiro, Itaara, Ivorá, Jari, Júlio de Castilhos, Nova Palma, Pinhal Grande, Quevedos, Santa Maria, São João do Polêsine, São Martinho da Serra, São Pedro do Sul, Silveira Martins, Toropi e Tupanciretã. Suas principais características:

- População Total (2018): 406.632 habitantes
- Área (2015): 12.395,9 km²
- Taxa de analfabetismo de pessoas com 15 anos ou mais (2010): 4,26 %
- Expectativa de Vida ao Nascer (2000): 72,82 anos
- Coeficiente de Mortalidade Infantil (2017): 8,92 por mil nascidos vivos
- Exportações Totais (2014): U\$ FOB 59.173.721

Figura 1 - Mapa do COREDE Central (2010)



Fonte: [Seplan/Deplan](#)

Segundo Kageyama (2004), já foi publicado pelo PNUD desde 1990 que fazer análise considerando apenas do Produto Interno Bruto per capita é insuficiente como medida de desenvolvimento. É necessário incorporar na análise outras variáveis, tendo em vista que a análise de desenvolvimento inclui o aumento de bem-estar da sociedade, que vai além da simplificação do aumento da renda.

Esta pesquisa vem com a intenção de dar continuidade ao estudo das desigualdades com uma ampliação para a construção do Índice de

Desenvolvimento Rural para o COREDE Central-RS, região onde está instalado o IFFar-JC – Instituto Federal Farroupilha – campus Júlio de Castilhos. A temática justifica-se por tratar-se de uma Instituição de ensino com cursos do setor agropecuário em uma região de perfil rural, sendo assim percebe-se a importância de um estudo nesse setor.

No próximo item passa-se a discorrer sobre análise multivariada com a finalidade de construir o referencial necessário para fundamentar a análise dos resultados.

2.3 ANÁLISE MULTIVARIADA

A análise multivariada é uma técnica estatística que é utilizada quando se tem uma grande quantidade de variáveis, as quais por meio da análise exploratória de dados é possível identificar e analisar a relação sistemática entre as variáveis de estudo. (VIALI, 2015)

De acordo com Hair et al (2009, pág. 23), a análise multivariada “se refere a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre indivíduos ou objetos sob investigação”.

A análise multivariada está sendo largamente aplicada na área das ciências sociais e aplicadas e se difundiu devido ao acesso aos softwares estatísticos que possibilitam ao pesquisador trabalhar com técnicas de análises sofisticadas. A configuração dos aplicativos viabiliza a criação de modelos quantitativos que tratam de previsão e explicação de fenômenos sociais e econômicos com capacidade para lidar com grande volume de dados de forma simultânea, identificando relações e interações entre as variáveis.

Existem dois grandes grupos para analisar variáveis: a estatística univariada que analisa as variáveis de forma isolada e a estatística multivariada que analisa as variáveis de forma conjunta, levando em conta todas as variáveis de forma simultânea, considerando as inter-relações que existem entre as variáveis (VICINI et al, 2018).

A análise multivariada aborda um conjunto de técnicas, as quais neste estudo considerou-se algumas delas que são: Análise de Agrupamento (*Cluster Analysis*), Análise Fatorial (*Factor Analysis*) e ACP – Análise de Componentes Principais.

2.3.1 Análise de Agrupamento (*Cluster*)

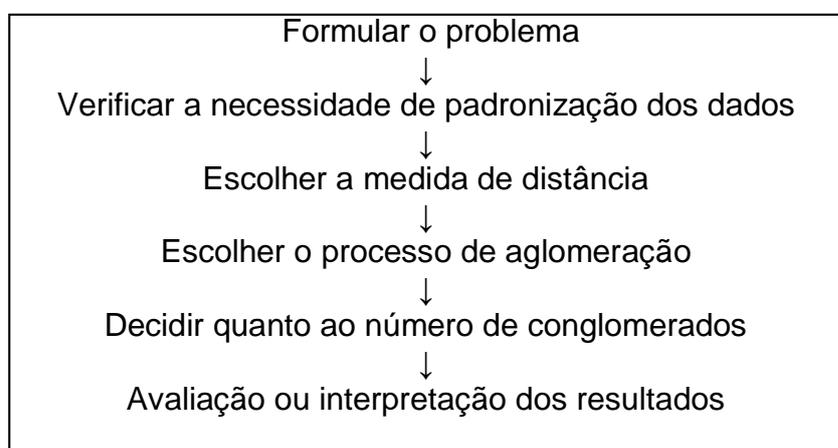
A análise de agrupamento é uma técnica multivariada utilizada para agregar objetos ou indivíduos considerando suas características comuns sem perder informações de todo o conjunto de estudos, de modo que cada objeto tenha semelhança aos outros, formando grupos homogêneos intragrupo e heterogêneos intergrupos (VICINI et al, 2018).

O primeiro passo para realizar a AA consiste em formular o problema de aglomeração, definindo as variáveis sobre as quais se baseará o agrupamento. Logo após, faz-se a coleta dos dados, que serão reunidos numa tabela com p colunas (variáveis) e n linhas (objetos) (VICINI et al, 2018, p. 39).

De acordo com Hair *et al* (2009) em cada cluster, os objetos tendem a ser similares entre si, mas são diferentes dos objetos de outros clusters. Segundo o autor após ser escolhida e calculada a medida de similaridade, deve-se formar os agrupamentos com base na medida de similaridade de cada observação. Existem dois procedimentos para a formação de agrupamentos: hierárquico e não hierárquico, sendo a aglomeração hierárquica a mais utilizada, pois, segundo Vicini *et al* (2018), ela faz a divisão de g grupos sem ser preciso definir a quantidade de grupos que os elementos de estudos serão subdivididos. Ela caracteriza-se por “uma estrutura em forma de árvore”.

O gráfico que representa o histórico do processo de aglomeração é denominado dendograma, o qual demonstra a síntese do trabalho desenvolvido, pois permite a classificação, comparação e discussão dos agrupamentos das variáveis.

Quadro 1 - Procedimentos para análise de agrupamento (AA)



Fonte: Elaborado pelo autor

As medidas de similaridade com foco na proximidade são baseadas em distância, as quais representam parecença com a proximidade de observações das variáveis agrupadas HAIR *et al* (2009). A distância euclidiana foi a medida de distância utilizada nesta pesquisa que, de acordo com Vicini *et al* (2018), considerando dois indivíduos i e i' , a distância entre eles é dada pela equação 1:

$$d_{ii'} = \left[\sum_{j=1}^p (X_{ij} - X_{i'j})^2 \right]^{\frac{1}{2}} \quad (1)$$

Onde:

$p = 1, 2, 3, \dots, j$;

X_{ij} = valor da variável j para o indivíduo i ;

$X_{i'j}$ = valor da variável j para o indivíduo i' .

O procedimento de encadeamento hierarquico de ligação utilizado é o método de Ward, que, segundo Hair *et al* (2009, pág. 452), “a similaridade entre dois agrupamentos não é a única medida de similaridade, mas a soma dos quadrados dentro dos agrupamentos feita sobre todas as variáveis”. O autor afirma que esse procedimento combina agrupamentos com um número de pequeno de observações, devido ao fato de que a soma dos quadrados está relacionada com o número de variáveis envolvidas. Esse método busca minimizar o quadrado da distância euclidiana e as médias dos *clusters*. A função é dada por Ward (1963), na equação 2.

$$d_{(UV)W} = \frac{(N_W + N_U) \cdot d_{UW} + (N_W + N_V) \cdot d_{VW} - N_W \cdot d_{UV}}{N_W + N_U + N_V}$$

Onde:

N_U e N_V são os números de elementos no grupo U e V, respectivamente.

d_{UW} e d_{VW} são as distâncias entre os elementos UW e VW, respectivamente.

A análise de agrupamento é um conjunto de técnicas multivariada que tem como meta reunir objetos de acordo com suas características, sendo bem sucedido se os objetos dentro do agrupamento ficam próximos uns dos outros

quando representados geometricamente, e quando aos agrupamentos diferentes estarão distantes entre si (HAIR *et al*, 2009). Destaca-se que o procedimento hierarquico fornece um excelente referencial para se comparar conjuntos de soluções de agrupamentos.

2.3.2 Análise Fatorial

O surgimento da análise fatorial ocorreu no século XX, com Karl Pearson, em 1901 e Charles Spearman, em 1904. Em 1947 Thurstone aproveitou os estudos iniciais desenvolvidos por Spearman e criou a análise fatorial. (FIGUEIREDO FILHO; SILVA JUNIOR, 2010)

Essa técnica é uma forma de análise multivariada que possibilita ao pesquisador trabalhar com grande número de variáveis e seus fatores subjacentes e analisar suas relações para explicar essas variáveis em termos de suas dimensões inerentes comuns que são os fatores. (MALHOTRA, 2012) Procura analisar um número menor de variáveis (fatores) sem que haja perda significativa de informação. (HAIR *et al*, 2009)

De forma simplificada, Vicini *et al* (2018, p. 150) afirmam que:

A análise fatorial espera agrupar variáveis originais em subconjuntos de novas variáveis independentes, e o que o pesquisador procura descobrir são os motivos que levaram a esse agrupamento. [...] Esse método determina quais variáveis pertencem a quais fatores, e o quanto cada variável explica cada fator. [...] utilizar a técnica de AF permitirá concluir se é possível explicar esse padrão de correlações mediante um menor número de variáveis.

Na visão de Malhotra (2012), a análise fatorial denota uma classe de processos utilizados para redução e sumarização dos dados e sua utilização visa identificar:

- fatores que expliquem as correlações entre um conjunto de variáveis;
- conjunto de variáveis não correlacionadas que substitui o conjunto original das variáveis correlacionadas, os quais poderiam ser regressão ou análise determinante;
- conjunto menor de variáveis que se destacam para uso de uma análise multivariada subsequente como ACP – análise de componentes principais.

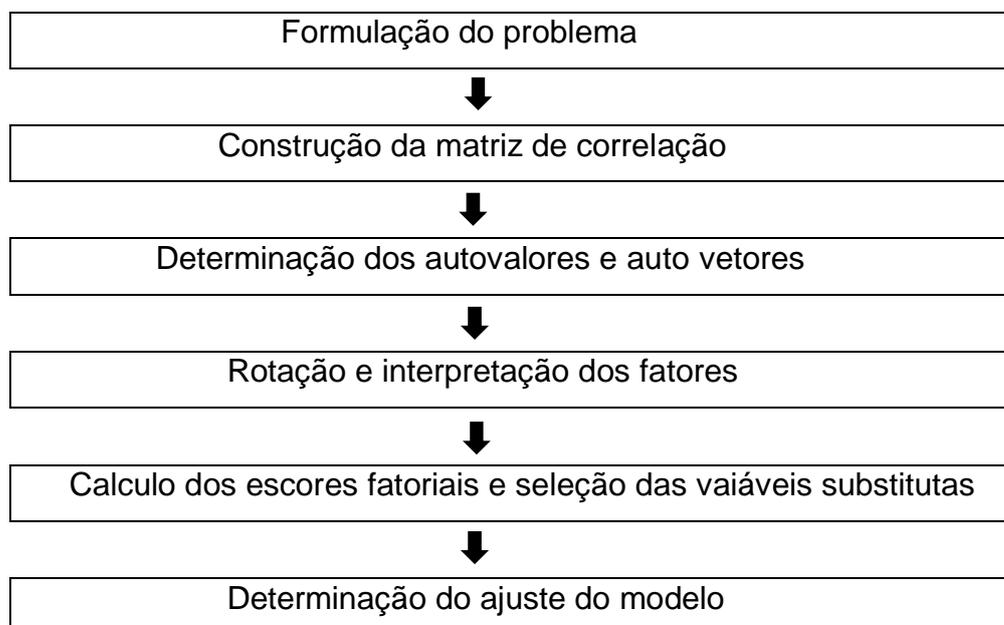
O objetivo da AF é demonstrar as relações de covariância entre as variáveis, supondo que elas possam formar um combinação de poucos fatores

independentes entre si. Os fatores são reconhecidos como variáveis latentes, constructo ou dimensões. (LOESCH; HOELTGEBAUM, 2012), (VICINI et al, 2018).

Dentro desse contexto, Malhotra (2012) afirma que a AF também é conhecida como análise fatorial exploratória, e se ajusta de acordo com quatro questões: redução dos dados; seleção das variáveis; utilização dos resultados por outras técnicas multivariadas.

A AF mostra aos pesquisadores quais variáveis podem se relacionar em conjunto e quantas destas podem ter impacto na análise (HAIR *et al*, 2009). Na Figura 1 apresentam-se os passos para a realização da AF:

Figura 2 – Passos para realizar a análise fatorial



Fonte: Adaptado de Mingotti (2007)

Para Malhotra (2012) na formulação do problema busca-se definir o objetivo da AF e posteriormente especificar as variáveis considerando que o tamanho da amostra deve ser cinco vezes maior que as observações. Com relação a construção da matriz de correlação as variáveis devem ser correlacionadas, onde a comprovação se dá pela correlação significativa entre as variáveis.

A rotação dos fatores se torna necessária na medida em que faz com que a matriz fatorial seja mais simples, proporcionando uma melhor interpretação dos

fatores. De acordo com Mingotti (2007) ao rotacionar os fatores serve para fazer com que cada fator tenha cargas e coeficientes não nulos, para algumas variáveis, e que cada variáveis tenha carregamentos significativos, com uns poucos fatores, sendo que a variável explicada pelos fatores individuais é redistribuída por rotação.

Os métodos de rotação mais utilizados são:

- **ortogonal** – rotação de fatores em que os eixos são mantidos em ângulo reto;
- **varimax** – método ortogonal de rotação de fatores que minimiza o número de variáveis com altas cargas em um fator, melhorando a interpretação das variáveis. É a mais utilizada e avalia as variâncias das cargas dentro de cada fator.

A AF, conforme Malhotra (2012); Vicini *et al* (2018); Hair *et al* (2009), tem o objetivo de reduzir e sumarizar o número de variáveis iniciais perdendo o menos possível de informação.

Um modelo de análise fatorial, de acordo com Mingoti (2007), é dado, genericamente, em forma matricial da seguinte forma:

$$X_i = a_{ij}F_j + \varepsilon_i \quad (1)$$

em que:

- $X_i = (X_1, X_2, \dots, X_p)^t$ é um vetor transposto de variáveis aleatórias observáveis;
- a_{ij} = é uma matriz (p x m) de coeficientes fixos denominados cargas fatoriais, os quais descrevem o relacionamento linear de X_i e F_j ;
- $F_j = (F_1, F_2, \dots, F_p)^t$ é um vetor transposto (m<p) de variáveis latentes que descrevem os elementos não observáveis da amostra;
- $\varepsilon_i = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)^t$ é uma vetor transposto dos erros aleatórios, correspondentes aos erros de medição e à variação de X_i que não é explicada pelos fatores comuns F_j .

Um dos métodos mais conhecidos para extração dos fatores é a análise dos componentes principais que passaremos a tratar no próximo item.

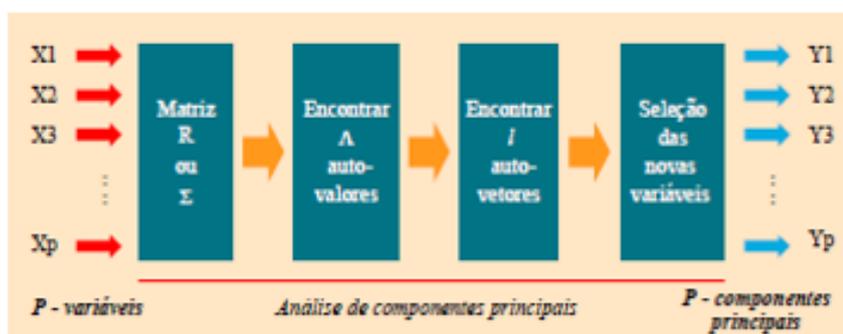
2.3.3 Análise de Componentes Principais – ACP

A análise de componentes principais propõe reduzir um conjunto de dados escritos em combinações lineares que foram determinados pelo autovalores e autovetores, estes estimados pela matriz de variância-covariância ou correlação, geralmente quando se trabalha com grande número de variáveis inter-relacionadas. (VICINI *et al*, 2018)

A ACP é uma técnica onde um conjunto de variáveis estudadas transforma-se em um novo conjunto de variáveis, no qual a variabilidade do conjunto original é mantida ao máximo. Essas novas variáveis são os Componentes Principais e são independentes e não correlacionados, favorecendo assim análises futuras. Segundo Hair *et al* (2009, pag. 33) "o objetivo da ACP é encontrar um meio de condensar a informação contida em várias variáveis originais em um conjunto menor de informação."

Para determinar os componentes principais, de acordo com Vicini *et al* (2018) é preciso calcular a matriz de variância-covariância ou a matriz de correlação, após encontram-se os autovalores e autorvetores e no final escreve-se as combinações lineares que serão as novas variáveis, os quais serão os componentes principais, conforme pode visualizar na Figura 2.

Figura 3 – Esquema para análise de componentes principais



Fonte: VICINI *et al* (2018, p. 88)

A demonstração genérica do modelo de ACP será feita de acordo com Hongyu, Sandanielo, Junior (2016), onde se tem as variáveis X_1, X_2, \dots, X_p em cada um de n indivíduos ou unidades experimentais. Este conjunto de $n \times p$ medidas origina uma matriz de dados X ($n \times p$):

$$\mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix} \quad (1)$$

Os componentes principais de uma forma geral são obtidos por meio de um conjunto de p variáveis X_1, X_2, \dots, X_p com médias $\mu_1, \mu_2, \dots, \mu_p$ e variância $\sigma_1^2, \sigma_2^2, \sigma_3^2, \dots, \sigma_p^2$, respectivamente. Estas variáveis não são independentes e portanto, possuem covariância entre a i -ésima e k -ésima variável definida por σ_{ik} , para $i \neq k = 1, 2, \dots, p$. Assim as p variáveis podem ser expressas na forma vetorial por: $\mathbf{X} = [X_1, X_2, \dots, X_p]'$, com vetor de médias $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]'$ e matriz de covariância $\boldsymbol{\Sigma}$ (HONGYU, SANDANIELO, JUNIOR, 2016).

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1p}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix} \quad (2)$$

Encontram-se os pares de autovalores e autovetores $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$, em que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, associados a $\boldsymbol{\Sigma}$ e então o i -ésimo componente principal é definido por (JOHNSON; WICHERN, 1998; HONGYU, 2015):

$$Z_i = e_i' \mathbf{X} = e_{i1} X_1 + e_{i2} X_2 + \cdots + e_{ip} X_p$$

A variável Z_i , é uma variável latente, ou seja, não é determinada a partir do experimento ou levantamento amostral (JOHNSON; WICHERN, 1998). O objetivo é mensura-la a partir das p variáveis contidas no vetor \mathbf{X} . A intenção é projetar os pontos coordenados originais em um plano maximizando a distância entre eles, o que equivale a maximizar a variabilidade da variável latente Z_i . A variância de Z_i é dada por

$$\text{Var}(Z_i) = \text{Var}(e_i' \mathbf{X}) = e_i' \text{Var}(\mathbf{X}) e_i = e_i' \boldsymbol{\Sigma} e_i \quad \text{em que } i = 1, \dots, p.$$

Os autores explicam e demonstram que utilizando a decomposição espectral da matriz $\boldsymbol{\Sigma}$, dada por $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$, em que \mathbf{P} é a matriz composta pelos

autovetores de Σ em suas colunas e Λ , a matriz diagonal de autovalores de Σ , então, tem-se que

$$tr(\Sigma) = tr(\mathbf{P}\Lambda\mathbf{P}') = tr(\Lambda\mathbf{P}'\mathbf{P}) = tr(\Lambda\mathbf{I}) = tr(\Lambda) = \sum_{i=1}^p \lambda_i$$

e

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix}$$

E $tr(\Sigma)$ é dada pela soma dos elementos da diagonal:

$$tr(\Sigma) = \sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i$$

Portanto, a variabilidade total contida nas variáveis originais é igual a variabilidade total contida nos componentes principais (JOHNSON; WICHERN, 1998). Ainda para Hongyu, Sandanielo, Junior (2016). A contribuição de cada componente principal (Z_i) é expressa em porcentagem, e a explicação individual de cada componente pode ser calculada, por exemplo, para késimo componente principal a proporção da explicação é dada por:

$$Ck = \frac{Var(Z_i)}{\sum_{i=1}^p Var(Z_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100 = \frac{\lambda_i}{tr(S)} \cdot 100$$

Pela proporção de explicação da variância total, que o modelo de k componentes principais é responsável, pode-se determinar o número de componentes que se deve reter. Em muitos casos, adota-se modelos que expliquem pelos menos 80% da variação total (JOHNSON; WICHERN, 1998). De maneira geral, escolhe-se o componente principal que possui a maior importância (o primeiro componente principal) como sendo aquele de maior variância (máx. λ_i), que explique o máximo de variabilidade dos dados, o segundo componente de maior importância, a seguir escolhe-se o que apresenta a segunda maior variância e assim sucessivamente, até o componente principal de menor importância (MANLY, 1986). De outro modo, os

últimos componentes principais identificarão relações lineares entre as variáveis originais próximo de constante (JOHNSON; WICHERN, 1998; ANDERSON, 2003; FERREIRA, 2011).

O principal uso da ACP, no início, foi a redução da dimensionalidade, mas depois passou a ter utilidade na identificação de variáveis em conjunto de dados multivariados. Conforme Vicini et al (2018) a ACP é capaz de identificar a componente em cada estágio de processos produtivos, sendo possível uma investigação mais profunda no processo e na etapa na qual ocorreu a falta de controle. Também é possível identificar cada AC responsável por respectivo setor de economia, tais como a formação de índices de desenvolvimento, entre outros índices necessários aos estudos econômicos.

3. MATERIAIS E MÉTODOS

O capítulo apresenta a abordagem metodológica da pesquisa, a qual envolve múltiplas variáveis, sendo necessária a utilização da estatística multivariada para a análise dos dados. Para tanto, foi efetuada análise de agrupamentos, análise fatorial e análise dos componentes principais (ACP).

Com o objetivo de mensurar o desenvolvimento rural, foram analisadas 15 variáveis para 19 municípios pertencentes ao COREDE central/RS. Essas variáveis foram divididas em 4 dimensões: social, demográfica, econômica e ambiental para a construção do IDR – Índice de Desenvolvimento Rural.

3.1 CARACTERIZAÇÃO DA PESQUISA

Este estudo é caracterizado quanto à natureza como básica que gera conhecimentos novos úteis para o avanço da ciência sem finalidades imediatas, a ser utilizado em pesquisas aplicadas ou tecnológicas para aplicação de seus resultados, (PRODANOV; FREITAS, 2013). A forma de abordagem é quantitativa devido a análise estatística dos dados.

Quanto ao objetivo a pesquisa é considerada como exploratória, para Gil (2010) a pesquisa exploratória tem o propósito de desenvolver, esclarecer e modificar conceito e ideias, bem como promover uma maior familiaridade do pesquisador com o problema. Pode-se dizer que esse tipo de pesquisa tem como objetivo principal o aprimoramento de ideias ou a descoberta e considera os mais variados aspectos relativos ao fato estudado.

Com relação aos procedimentos técnicos trata-se de pesquisa bibliográfica, esta pesquisa é elaborada com base em materiais já publicados, tradicionalmente, inclui material impresso, dissertações e anais de eventos científicos (GIL, 2010), e documental, é chamada assim, porque procura em documentos de fonte primária e secundária, a saber, dados provenientes de órgãos que realizam as observações, esses dados podem ser encontrados em jornais, arquivos e fontes estatísticas (RAMPAZZO, 2005). Carvalho (2008), cita que pesquisa documental é aquela realizada a partir de documentos considerados legítimos, a fim de, comparar fatos sociais, estabelecendo características de uma realidade.

A consulta e extração dos dados foram feitas em documentos oficiais do IBGE (Instituto Brasileiro de Geografia e Estatística) e da FEE (Fundação de Economia e Estatística).

Dentro deste contexto, o que os autores anteriormente citados expõem sobre os conceitos acerca dos tipos de metodologia e suas aplicações, foi aplicado neste trabalho a pesquisa exploratória, com embasamento bibliográfico e com coleta documental e análise quantitativa.

Quadro 2- Classificação e enquadramento metodológico

CLASSIFICAÇÃO	ENQUADRAMENTO	AUTORES
Natureza	Básica	PRODANOV; FREITAS (2013)
Forma de abordagem	Quantitativa	GIL (2010)
Objetivos	Exploratória	GIL (2010)
Procedimentos técnicos	Bibliográfica Documental Estatística	CARVALHO (2008) GIL (2010) RAMPAZZO (2005)

Fonte: Elaborado pelo autor.

3.2 ÁREA DE ESTUDOS E VARIÁVEIS ANALISADAS

A área de estudo considerada foi o COREDE Central, o qual é composto por 19 municípios pertencentes a microrregião Central do Rio Grande do Sul: Agudo, Dilermando de Aguiar, Dona Francisca, Faxinal do Soturno, Formigueiro, Itaara, Ivorá, Jari, Júlio de Castilhos, Nova Palma, Pinhal Grande, Quevedos, Santa Maria, São João do Polêsine, São Martinho da Serra, São Pedro do Sul, Silveira Martins, Toropi e Tupanciretã.

As variáveis consideradas para construir o IDR – Índice de Desenvolvimento Rural, são pertencentes a quatro dimensões: social, demográfica, econômica e ambiental, conforme a seguir:

Quadro 3 – Variáveis de desenvolvimento rural utilizadas – ano 2010

VARIÁVEIS	DESCRIÇÃO	DIMENSÃO
ERW	Consumo de energia elétrica rural por mhw (Megawatt-hora)	Social
ECO	Consumo de energia elétrica rural por nº de consumidores	Social
AGU	Nº domicílios rurais com rede geral de abastecimento de água	Social
LIX	Domicílios rurais com coleta de lixo	Social
POP	População rural	Demográfica
DOM	Domicílios particulares permanentes na zona rural	Demográfica
POC	Pessoas ocupadas no ambiente rural	Demográfica
ARE	Área de estabelecimentos rurais	Demográfica
SOJ	Valor da produção de soja	Econômica
ARR	Valor da produção de arroz	Econômica
VAB	Valor adicionado bruto na agropecuária	Econômica
APA	Área plantada de arroz (ha)	Ambiental
APS	Área plantada de soja (ha)	Ambiental
ECP	Nº de estabelecimentos agropecuários com controle de pragas e/ou doenças vegetais	Ambiental
EPS	Nº de estabelecimentos agropecuários com sistema de preparo do solo	Ambiental

Fonte: PINTO E CORONEL (2016, p. 905)

3.3 MÉTODO DE ANÁLISE

O presente estudo baseia-se em trabalhos já existentes na literatura que fizeram uso de uma metodologia para criar um IDR – Índice de Desenvolvimento Rural. A partir desses estudos relativos ao tema, percebe-se que o fenômeno de desenvolvimento rural possui um caráter multidimensional, isto é, que a magnitude do problema requer a consideração de um conjunto de variáveis de características locais diversas (CONTERATO; SCHNEIDER; WAQUIL, 2007; MELO; PARRÉ, 2007)

De acordo com essa ideia, devido ao grande número de variáveis consideradas, foi necessária a utilização da técnica estatística de análise multivariada, sendo a análise de cluster, a análise fatorial e a análise dos componentes principais as ferramentas mais adequadas para este estudo (ALVES, 2012; MELO; PARRÉ, 2007; STEGE; PARRÉ, 2011).

O modelo de análise fatorial de acordo com Mingotti (2007), em forma matricial, foi demonstrado de forma genérica no capítulo do referencial teórico deste pesquisa.

Devido ao fato de as variáveis componentes do índice possuírem unidade de valores diferentes, é necessário padronizar as variáveis, a fim de tornar os objetos de estudo comparáveis a partir da diminuição das escalas diferentes. (BASSAB; MIAZAKI; ANDRADE, 1990). A padronização das variáveis é dado pelo seguinte procedimento:

$$Z = \frac{(X_i - \bar{X})}{S}, \quad i=1, \dots, n \quad (2)$$

em que:

Z = variável padronizada;

X_i = variável a ser padronizada;

\bar{X} = média de todas as observações;

S = desvio padrão amostral.

Após padronizar as variáveis aleatórias observáveis, serão substituídas pelas variáveis padronizadas Z_i , a fim de resolver o problema de diferenças de unidade de escala, como mostra a equação (2), (MINGOTI, 2007).

A análise fatorial é uma abordagem que busca analisar as correlações entre um grande grupo de variáveis, definindo um conjunto de dimensões latentes comuns, que são os fatores (HAIR et al, 2009).

Sendo assim, com a identificação dos fatores presentes entre as variáveis, há a possibilidade de determinar as relações quantitativas considerando aquelas que apresentam semelhança e definir os fatores componentes do desenvolvimento rural. Nesta pesquisa, foi aplicado o recurso de rotação dos fatores originais pelo método Varimax Normalizada, o qual faz com que os eixos sejam rotacionados permitindo uma melhor visualização da disposição dos pontos. Esse método demonstra uma estrutura simples de ser interpretada por que maximiza as correlações de cada variável em um único fator (HAIR et al, 2009).

No presente estudo foram agregadas as observações das variáveis componentes do IDR referente ao ano de 2010, com dados do IBGE e da FEE, por ser o ano no Censo, o qual é feito a cada dez anos.

Com o objetivo de verificar a adequação dos dados na análise fatorial utilizou-se o teste de Bartlett para verificar a significância e o Critério de *Kaiser-*

Meyer-Olkin (KMO). Este teste verifica a adequação dos dados a partir da criação de um índice que varia de 0 a 1, que compara as correlações simples e parciais entre as variáveis, sendo que valores superiores a 0,5 demonstram que os dados são adequados à análise fatorial (HAIR et al, 2009; MINGOTI, 2007). Além disso, para testar a confiabilidade dos dados utilizados para compor os fatores, foi estimado o Alfa de *Cronbach*.

Ao analisar as questões sociais, econômicas e ambientais percebe-se a dificuldade de mensuração por que envolve vários elementos correlacionados e verifica-se a necessidade de considerar a dinâmica do fenômeno e seus aspectos multidimensionais para as transformações dessas questões em números, ou seja, por meio de índices (CONTERATO; SCHNEIDER; WAQUIL, 2007). A partir disso, o Índice de Desenvolvimento Rural surge como uma escala de medida com o objetivo de mensurar o desenvolvimento rural de determinado território. Para o presente trabalho, os índices foram calculados para o ano 2010.

Sintetizando, neste trabalho, foi feita, primeiramente, uma análise de *Cluster* com a finalidade de agrupar as variáveis de acordo com sua semelhança. Após foi feita Análise Fatorial para a redução das variáveis por meio da formação dos fatores e posteriormente, foi feita a ACP – análise dos componentes principais para a formação do IDR, com os componentes selecionados.

Os dados estatísticos utilizados foram obtidos de fontes primárias e secundárias pela coleta de dados nos *sites* do IBGE e FEE. A partir disso, utilizou-se os softwares Statistica 9.1 para a análise dos resultados da pesquisa, demonstradas no próximo capítulo.

4. ANÁLISE DOS RESULTADOS

A finalidade deste capítulo é apresentar os resultados obtidos por meio do levantamento das 15 variáveis de desenvolvimento rural para o ano de 2010, nos 19 municípios do COREDE central/RS, o método de análise fatorial foi realizado a fim de indicar os fatores que determinam o desenvolvimento rural a partir dessas variáveis, anteriormente foi feita a análise de cluster para verificar a similaridade e selecionar as variáveis para a análise fatorial.

4.2 ANÁLISE DE CLUSTER

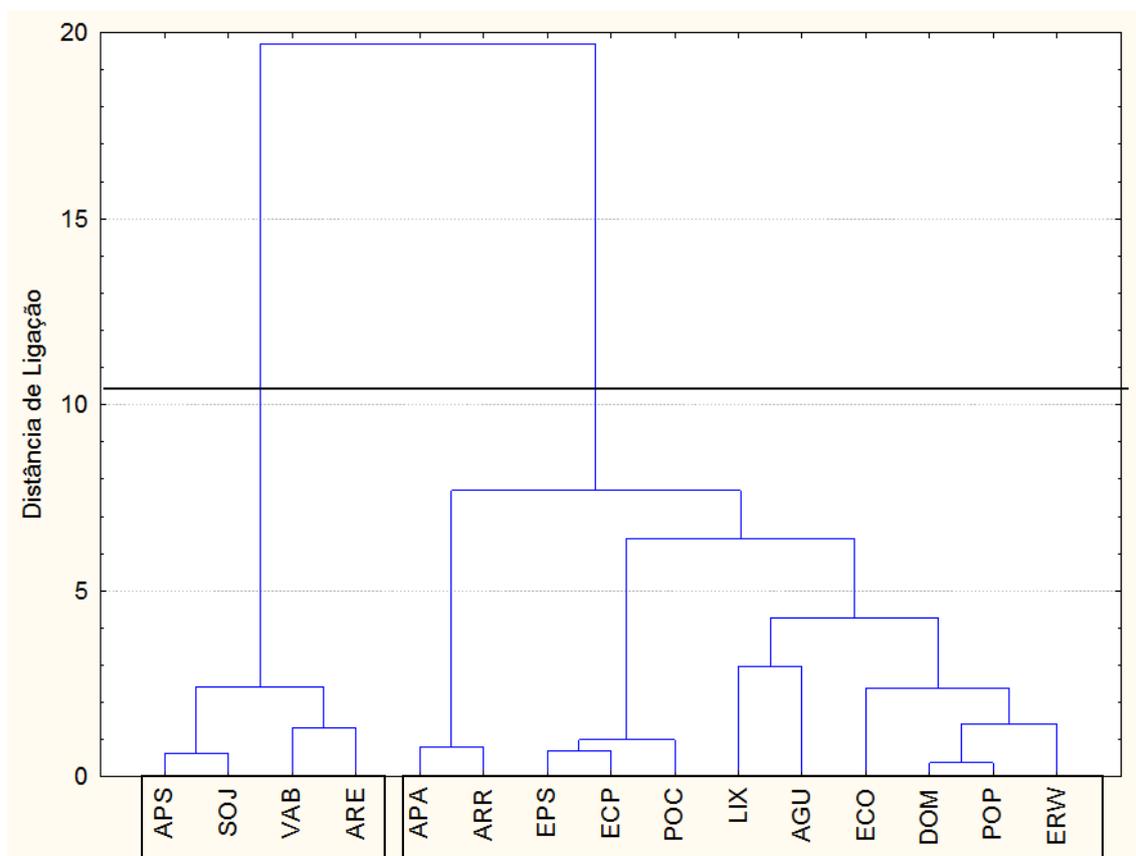
A utilização da análise de agrupamentos pressupõe alguns critérios, como a estrutura dos dados e o método considerado de agrupamento, pois o seu objetivo principal é definir a estrutura dos dados e agrupar as observações mais parecidas.

Por meio da análise de agrupamentos foi possível verificar a formação de vários clusters, onde a intenção é chegar a um número mais reduzido de variáveis sem ocorrer perda de informação, retirando as variáveis que apresentam similaridade. Considerou-se a ideia de que os objetos que apresentam distância euclidiana menor pertencem ao mesmo cluster (SOUZA; VICINI, 2005).

Na pesquisa foi utilizado o método de Ward como algoritmo de aglomeração e a distância euclidiana como medida de similaridade. Para identificar o número de clusters foi feito o ponto de corte no dendrograma demonstrando a maior distância entre as variáveis.

Na Figura 4, mostra-se o comportamento do dendrograma com todas as variáveis originais, no qual pode-se identificar a formação de dois *clusters*, obtidos por meio de um corte transversal feito na parte de maior distância entre os *clusters*. O primeiro *clusters* é formado pelas variáveis *APS, SOJ, VAB E ARE*, e o outro *cluster* formado pelas variáveis *APA, ARR, EPS, ECP, POC, LIX, AGU, ECO, DOM, POP E ERW*.

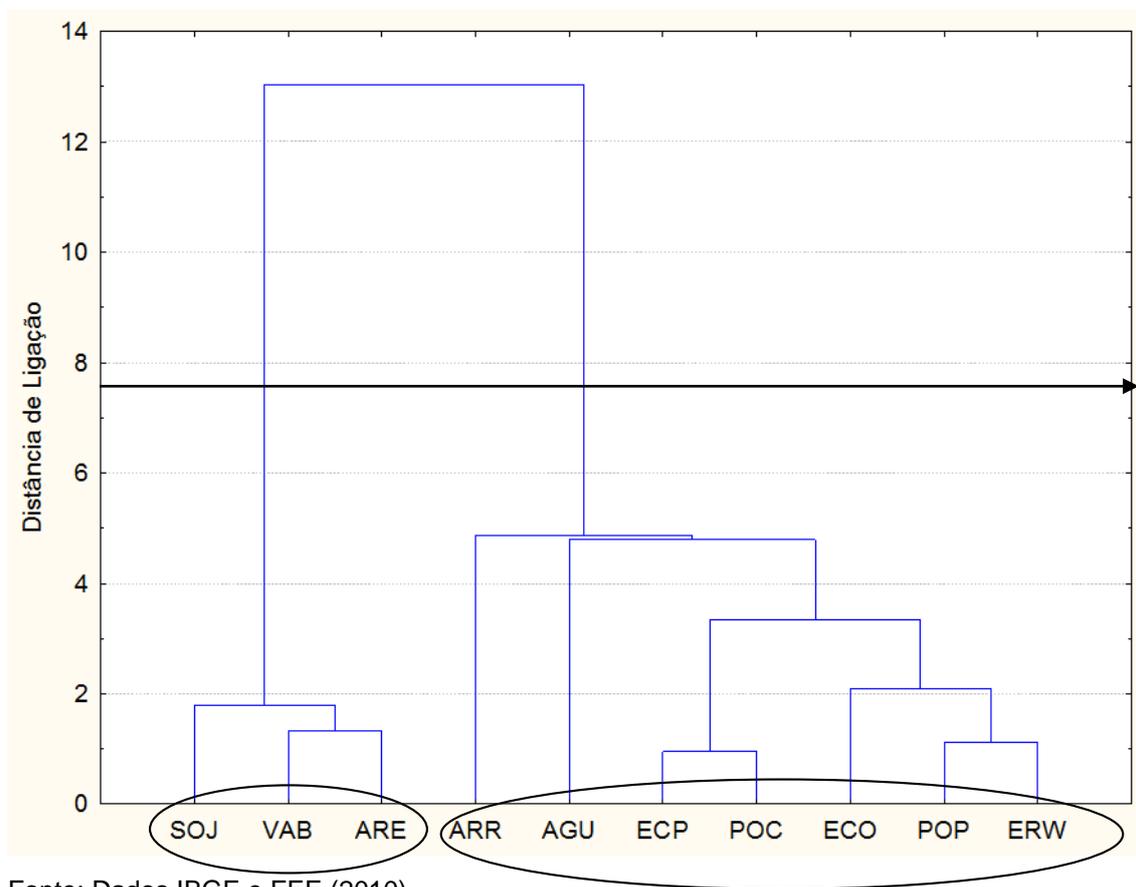
Figura 4 – As 15 variáveis relacionadas aos 19 municípios do COREDE Central



Fonte: Dados IBGE e FEE (2010)

Na análise do dendrograma, decidiu-se retirar as variáveis que possuem a mesma representação dentro do *cluster*, devido ao fato de pertencerem a mesma dimensão. No primeiro *cluster* as variáveis *APS* (*área plantada de soja*) e *SOJ* (*valor da produção de soja*) pertencem a dimensão econômica e apresentam similaridade, pois possuem a menor distância euclidiana, sendo assim retirou-se a variável *APS*. O segundo *cluster* retirou-se as variáveis *dom* (*domicílios particulares permanentes*), *EPS* (*número de estabelecimentos agropecuários com sistema de preparo do solo*), *APA* (*área plantada de arroz*) e *LIX* (*domicílios rurais com coleta de lixo*), pois elas pertencem, respectivamente, a mesma dimensão, não havendo portanto, perda de informação, visto que sua similaridade é comprovada por que possuem a menor distancia euclidiana. Forma-se um novo dendrograma na Figura 5, no qual pode-se verificar a formação de dois clusters. (SOUZA; VICINI, 2005)

Figura 5 – Dendrograma com 10 variáveis, após a retirada de variáveis com o mesmo grau de similaridade

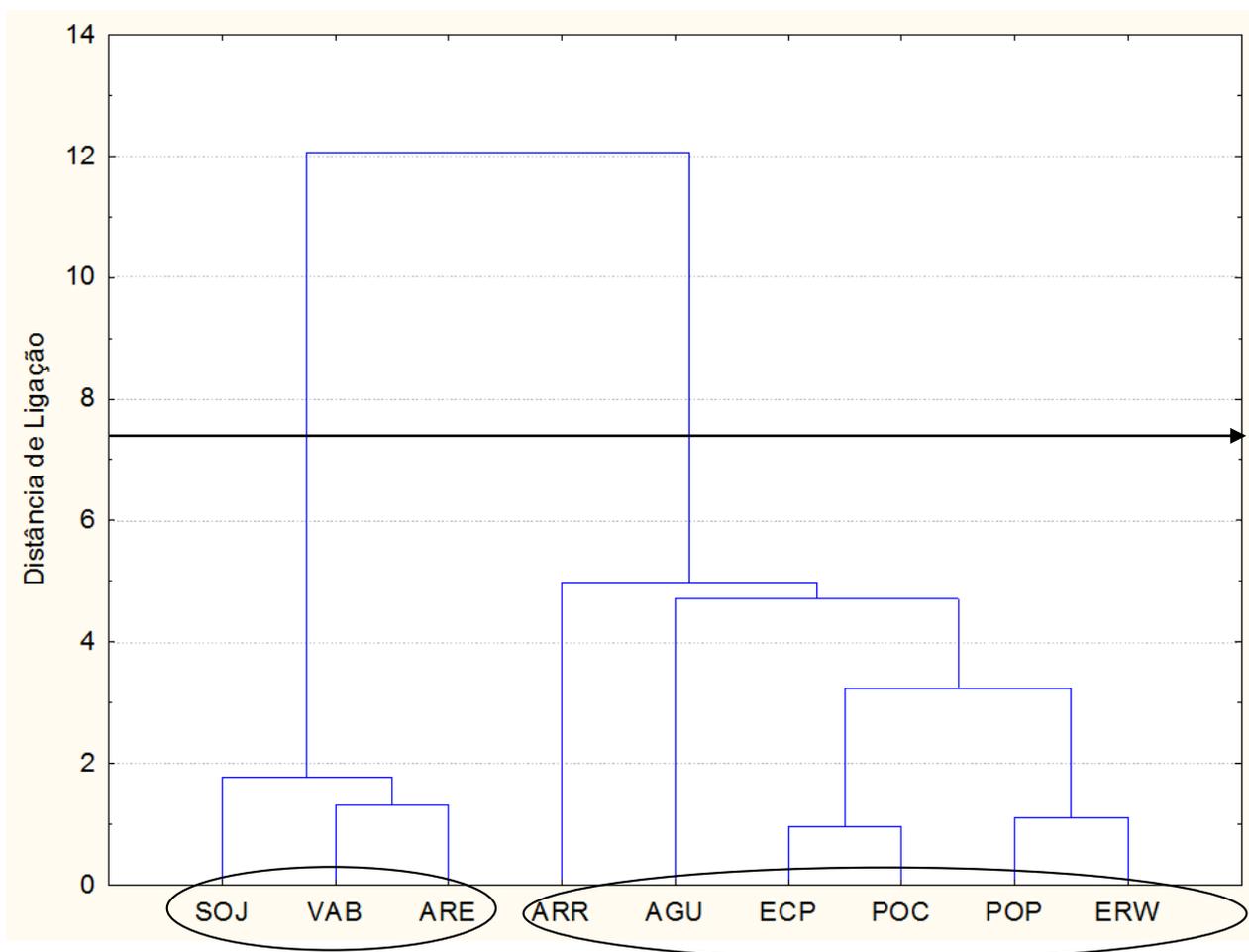


Fonte: Dados IBGE e FEE (2010)

Na segunda análise do dendrograma (Figura 5) retirou-se a variável *ECO* por fazer parte da mesma dimensão social que *ERW*, nesse sentido são variáveis similares que na redução não haverá perda de informação, e forma-se outro dendrograma (Figura 6) também com dois clusters.

Na terceira análise de agrupamentos foi possível apresentar o dendrograma final, que pode ser visualizado na Figura 6, pode-se verificar a formação de dois clusters, os quais possuem variáveis de maior relevância dentro do conjunto original das variáveis pesquisadas. (SOUZA; VICINI, 2005)

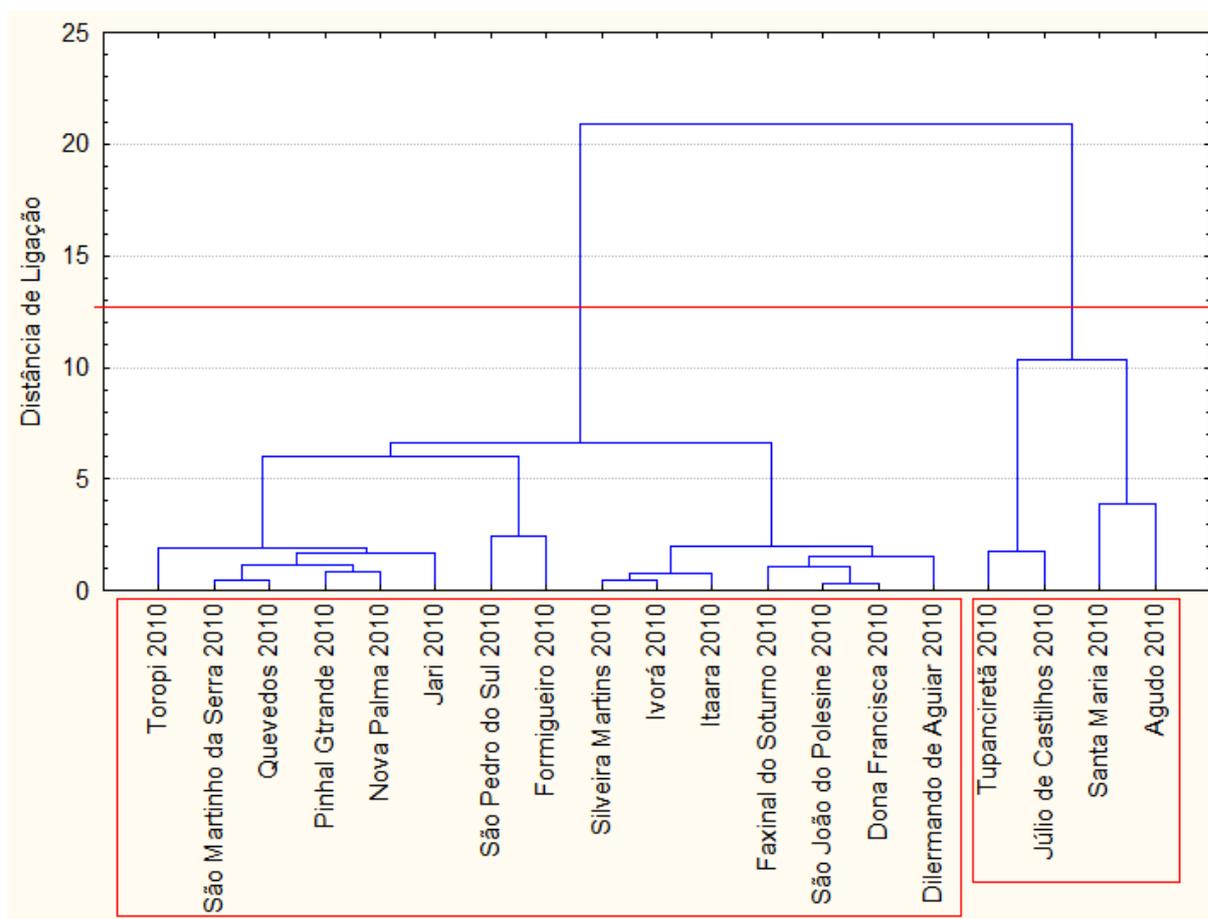
Figura 6 – Dendrograma com 9 variáveis, após a retirada de variáveis com o mesmo grau de similaridade



Fonte: Dados IBGE e FEE (2010)

Na figura 7 foi feita uma análise de agrupamento considerando os 19 municípios do COREDE Central, foi possível observar a formação de 2 clusters, onde 15 municípios apresentaram similaridade formando um cluster e os 4 municípios restantes concentraram no outro cluster.

Figura 7 - Dendrograma com os 19 municípios do COREDE Central



Fonte: Dados IBGE e FEE (2010)

Identificadas as variáveis como pertencentes ao mesmo *cluster* e permanecendo na análise as variáveis consideradas mais relevantes sob a ótica do pesquisador, procedeu-se a análise fatorial dessas variáveis (SOUZA; VICINI, 2005).

Na verdade, o que se percebe a partir da Figura 6 das variáveis é que os quatro municípios, Agudo, Júlio de Castilhos, Santa Maria e Tupanciretã apresentam similaridade em relação as variáveis SOJ, VAB e ARE. A Figura 7 mostra o agrupamento destes quatro municípios como predominantes na dimensão econômica de acordo com as variáveis acima, caracterizando o COREDE Central como o maior produtor de soja do estado.

4.3 ANÁLISE FATORIAL

A análise fatorial foi efetuada considerando as variáveis selecionadas na análise de cluster. Inicialmente eram 15 variáveis para 19 municípios do COREDE central/RS, pertencendo a 4 dimensões necessárias para verificar o desenvolvimento rural, sendo elas: social, demográfica, econômica e ambiental.

Após a análise de agrupamento selecionou-se 9 variáveis relevantes para o estudo, deste modo estas variáveis formam a matriz de observações, indicando as dimensões das variáveis estatísticas com perda mínima de informações. Com o propósito de verificar o peso das variáveis, realizou-se a análise fatorial extraíndo os fatores por meio da análise de componentes principais. Foram realizados os testes de Bartlett e KMO para verificar a adequabilidade da realização da análise fatorial.

Os resultados demonstram significância para o teste de Bartlett ao nível de 1%, indicando que há rejeição da hipótese nula de que a matriz de correlação é uma matriz identidade. O KMO apresentou, para a análise fatorial das variáveis de desenvolvimento rural, valor de 0,661, o qual é maior que o valor crítico e adequado ao emprego da análise fatorial de 0,5 (HAIR *et al*, 2009). Sendo assim, os dois testes demonstram que é viável a realização da análise fatorial para os objetivos da pesquisa. Foi estimado o Alfa de Crombach para verificar a confiabilidade das variáveis e foi obtido um valor de 0,747, demonstrando um valor satisfatório (HAIR *et al*, 2009).

Tabela 1 – Autovalores da matriz e variância explicada das correlações para as variáveis de desenvolvimento rural dos municípios do COREDE central/RS – ano 2010

FATOR	AUTOVALOR	VARIÂNCIA EXPLICADA PELO FATOR (%)	VARIÂNCIA ACUMULADA (%)
1	5,582	62,02	62,02
2	2,275	25,28	87,30

Fonte: Elaborado pelo autor

Ao aplicar a análise fatorial com a utilização do método de componentes principais e com o método de rotação ortogonal Varimax Normalizada, nas

variáveis de desenvolvimento rural, conforme mostra na tabela 1, selecionou-se os dois autovalores maiores que 1 que explicam 87,30% de variância total dos dados explicada do conjunto das 9 variáveis selecionadas na análise de cluster.

Após serem definidos os fatores a serem utilizados, as cargas fatoriais e as comunalidades associadas a cada variável são apresentadas na Tabela 2. As comunalidades, as quais correspondem a proporção de variabilidade de cada variável que é explicada pelos fatores, mostram-se satisfatórias quando apresentam valores próximos ou superiores a 0,5. As cargas fatoriais que apresentam valores superiores a 0,700 (valores em destaque) demonstram maior contribuição de uma variável na composição de um fator.

Os valores encontrados para as comunalidades das variáveis de desenvolvimento rural revelam que todas têm sua variabilidade captada pelos dois fatores. Analisando a Tabela 2, verifica-se que o Fator 1 é composto pelas variáveis *ERW*, *POC*, *POP* e *ECP*, que representam, respectivamente, as variáveis consumo de energia elétrica rural por mhw, pessoas ocupadas no ambiente rural, população rural e número de estabelecimentos com controle de pragas e doenças vegetais. Todas essas variáveis apresentam carga fatorial positiva na composição do fator e abordam características do mercado de trabalho, das condições da população rural e do meio ambiente (PINTO; CORONEL, 2016). Portanto, esse fator pode ser denominado Fator Pessoas e Ambiente Rural e explica 62% das correlações das variáveis.

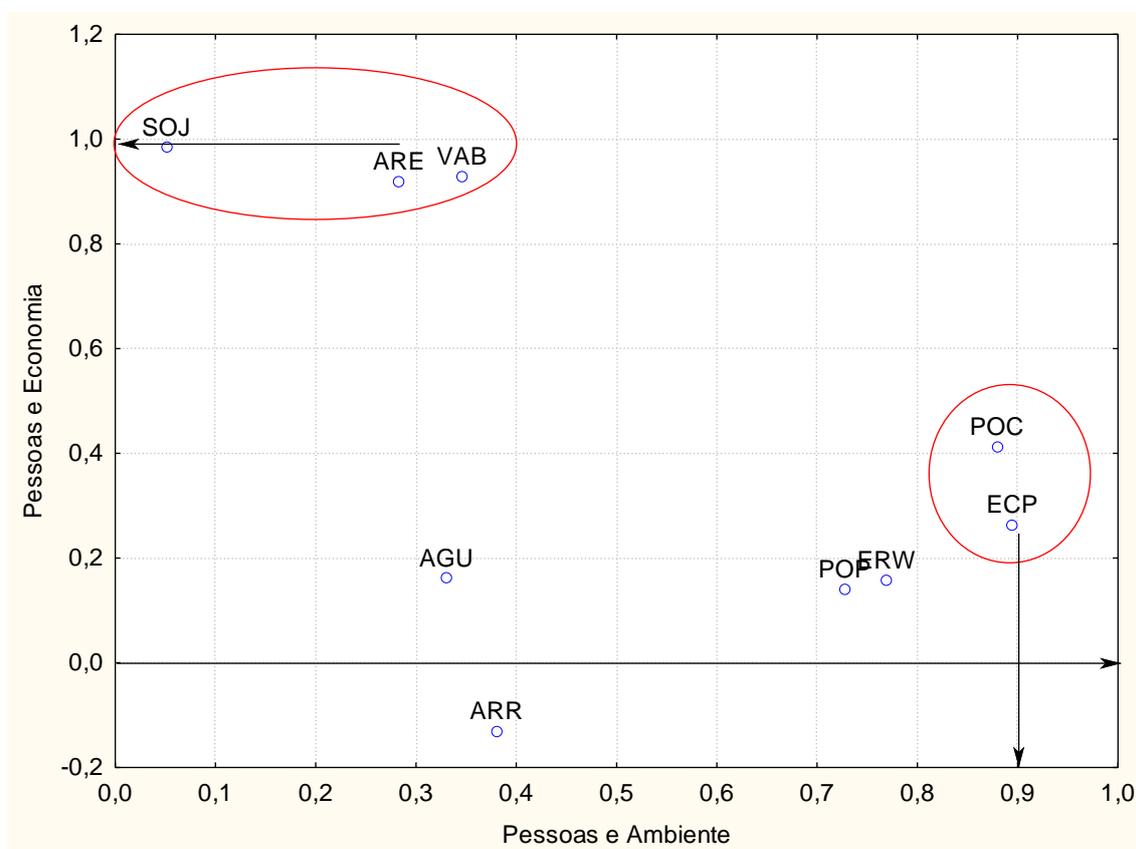
Tabela 2 – Cargas fatoriais das variáveis de desenvolvimento rural dos municípios do COREDE central/RS após rotação ortogonal e comunalidades – ano 2010

VARIÁVEIS	CARGAS FATORIAIS		COMUNALIDADES
	Factor 1	Factor 2	
ERW	0,7697	0,1576	0,6173
AGU	0,3298	0,1615	0,1348
POP	0,7276	0,1398	0,5489
ARE	0,2830	0,9190	0,9246
POC	0,8794	0,4116	0,9427
SOJ	0,0513	0,9841	0,9711
ARR	0,3803	-0,1309	0,1368
VAB	0,3459	0,9279	0,9806
ECP	0,8947	0,2638	0,8701

Fonte: Elaborada pelo autor

O Fator 2 possui, na sua composição, as variáveis *ARE*, *SOJ* e *VAB*, que, respectivamente, correspondem as variáveis: área dos estabelecimentos rurais, valor da produção de soja e valor adicionado bruto na agropecuária. As variáveis apresentam carga fatorial positiva na composição do fator e abordam aspectos quanto área e ao valor da produção da agropecuária do COREDE central/RS como um todo, bem como da cultura da soja, seu principal produto, (MELO; PARRÉ, 2007, PINTO; CORONEL, 2016), tendo em vista que os dois municípios que mais produzem soja no Rio Grande do Sul fazem parte do COREDE central, conforme reportagem da Gaúcha ZH (2018) “Tupanciretã lidera a lista de 20 municípios, com 496,3 mil toneladas produzidas de **soja**. Na sequência, Júlio de Castilhos apresenta a **maior** produção com 400,3 mil toneladas”⁴. Dessa forma, o Fator 2 pode ser denominado Pessoas e Economia. A seguir, na Figura 8, o gráfico que permite a visualização da relação do Fator 1 versus o Fator 2, com rotação Varimax normalizada.

Figura 8 - Gráfico que representa a relação entre fatores (Fator1 x Fator2)



⁴ Dados publicados em Gaúcha ZH – Campo e lavoura – 10/08/2018.

A Figura 8 demonstra a relação entre as variáveis denominadas de Pessoas⁵ e Ambiente (pessoas ocupadas no ambiente rural e número de estabelecimentos agropecuários com controle de pragas e doenças) com as denominadas de Pessoas e Economia (valor da produção de soja, valor agregado bruto no setor rural e área dos estabelecimentos rurais). Percebe-se, na análise, que as variáveis formam grupos por semelhança de explicação. As variáveis *POC* e *ECP* são as que melhor explicam o Fator 1 – Pessoas e ambiente por que são as mais distantes da origem. As variáveis que melhor representam o Fator 2 – Pessoas e economia são *SOJ*, *VAB* e *ARE*. As demais variáveis possuem baixa representatividade por estarem mais próximas da origem. (VICINI *et al*, 2018).

As variáveis do Fator 1 evidenciam a questão das pessoas ocupadas no setor rural (*POC*) em relação ao total da população rural e o número de estabelecimentos rurais com controle de pragas e/ou doenças vegetais (*ECP*), que caracteriza a questão da estrutura e desempenho do setor rural com relação ao meio ambiente, quanto ao Fator 2, é explicado pelas variáveis do valor da produção da soja (*SOJ*), pelo valor adicionado da produção (*VAB*) e área dos estabelecimentos rurais (*ARE*) demonstrando a evidencia da cultura da soja como principal formação de renda na região de estudo. Essa análise vem corroborar com as análises já feitas anteriormente, desde a análise de agrupamentos, a qual originou 2 clusters com essas variáveis.

4.4 A CONSTRUÇÃO DO IDR VIA ANÁLISE DE COMPONENTES PRINCIPAIS

A proposta deste trabalho é a construção do Índice de Desenvolvimento Rural (IDR) para o COREDE central/RS por meio da análise de componentes principais (ACP).

Aplicando ACP ao conjunto de 9 variáveis, resultaram dois fatores que representam 87,30% das variáveis originais. Dessa maneira, constatou-se que a primeira componente principal (CP1) representa 62,02% e a segunda (CP2) representa 25,28% da variabilidade das variáveis selecionadas, conforme já especificado na AF.

⁵ A escolha dos nomes Pessoas e Ambiente e Pessoas e Economia foi devido ao fato de as variáveis envolvidas serem pertencentes as dimensões demográfica, econômica e ambiental.

Tabela 3 - Contribuição das variáveis baseada nas correlações

VARIÁVEIS	CP1	CP2
	PESSOAS E AMBIENTE	PESSOAS E ECONOMIA
ERW	0,146	0,058
AGU	0,087	0,015
POP	0,147	0,068
ARE	0,109	0,147
POC	0,159	0,0001
SOJ	0,047	0,317
ARR	0,049	0,194
VAB	0,101	0,182
ECP	0,154	0,017

Fonte: Elaborado pelo autor

Para a construção do IDR foi considerado a contribuição da variáveis em cada Fator 1 (CP1) e Fator 2 (CP2), conforme demonstra a Tabela 3, o IDR1 corresponde a Pessoas e ambiente e o IDR2 a Pessoas e economia e estão demonstrados na Tabela 4.

4.4.1 Índices de Desenvolvimento Rural

$$\text{IDR 1} = 0,146.ERW + 0,087.AGU + 0,147.POP + 0,109.ARE + 0,159.POC + 0,047.SOJ + 0,049.ARR + 0,101.VAB + 0,154.ECP$$

$$\text{IDR 2} = 0,058.ERW + 0,015.AGU + 0,068.POP + 0,147.ARE + 0,0001.POC + 0,317.SOJ + 0,194.ARR + 0,182.VAB + 0,017.ECP$$

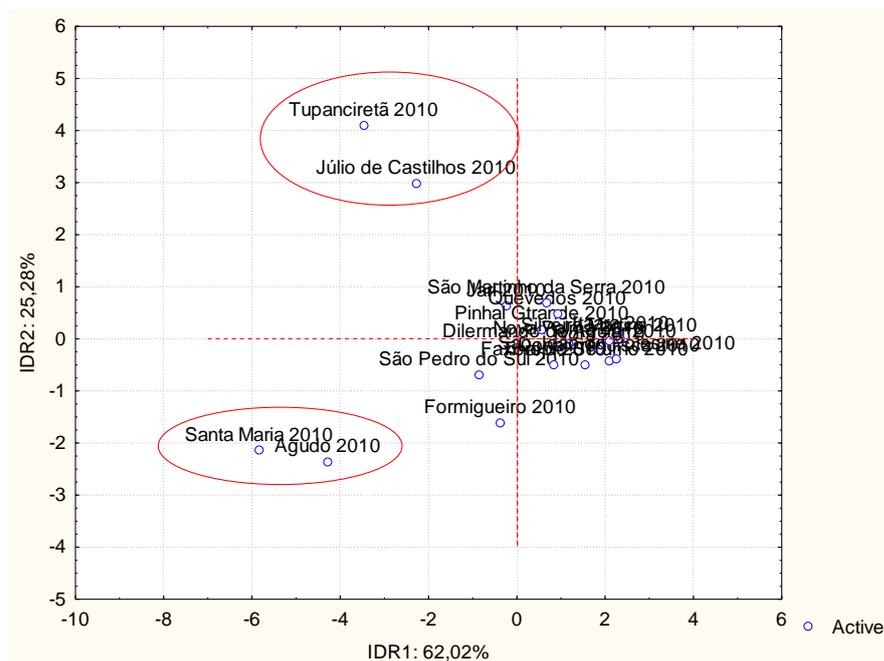
No item 4.4.1 está demonstrada a construção do IDR1 do COREDE Central/RS para Pessoas e Ambiente, onde verificou-se a maior contribuição das seguinte variáveis: População Rural (*POP*), Pessoas ocupadas no ambiente rural (*POC*) e Número de estabelecimento agropecuários com controle de pragas e doenças vegetais (*ECP*). O destaque das variáveis *POP* e *POC* que relaciona a população rural ocupada e o tratamento do solo (*ECP*) demonstra que o COREDE central/RS possui grande dependência das rendas geradas da atividade agropecuária (ALONSO, 2003). Percebe-se que essa região se desenvolve com predominância nas questões de estrutura e desempenho do

setor agropecuário, gerando emprego e rendas que proporcionem o desenvolvimento de todo o COREDE central/RS.

As variáveis de maior peso para a construção do IDR2 do COREDE CENTRAL/RS onde predominou Pessoas e Economia foram: Valor da produção de soja (SOJ), Valor da produção de arroz (ARR) e Valor adicionado bruto na agropecuária (VAB). Corroborando com o IDR1, os municípios do COREDE central/RS fundamentam-se na atividade rural. Essa região se desenvolveu com a introdução de novos produtos para o cultivo nas áreas de produção agropecuária, com predominância da soja (CONTERATO, 2008).

As variáveis que mais contribuem para a formação do IDR2 demonstram a elevada produção da soja e o arroz em segundo lugar, com o Valor Adicionado Bruto na agropecuária sendo destaque em relação aos demais setores da economia. Sendo comprova que o setor de alavancagem do COREDE central/RS é o setor rural por meio da cultura da soja.

Figura 9 – Gráfico de distribuição da nuvem de pontos (os municípios)



Fonte: Elaborado pelo autor

Observando a Figura 9 pode-se concluir que os municípios que mais contribuíram na formação da combinação linear do IDR1 foram Agudo e Santa Maria, eles representam as variáveis do Fator 1 (*POC*, *POP* e *ECP*). Pode-se concluir ainda que os municípios que possuem uma maior contribuição na formação do IDR2 são Júlio de Castilhos e Tupanciretã, que representam as

variáveis do Fator 2 (*SOJ*, *VAB* e *ARR*). O restante dos municípios e variáveis não apresentam representatividade significativa em relação a esses dois índices (VICINI *et al*, 2018).

Segundo dados do IBGE⁶ (2019) o maior valor da produção de soja está concentrado nos municípios de Júlio de Castilhos e Tupanciretã e também o valor da produção de arroz é nos municípios de Agudo, Santa Maria e Formigueiro, demonstrando assim que o desenvolvimento rural se apresenta de maneira heterogênea dentro dos municípios do COREDE central/RS.

⁶ Instituto Brasileiro de Geografia e Estatística

5. CONSIDERAÇÕES FINAIS

Com base nos resultados obtidos pode-se concluir que o desenvolvimento rural no COREDE central/RS, tendo em vista que a questão regional tem estado em pauta nas mais diversas instancias, no estado do Rio Grande do Sul, bem como tem sido tema nas pesquisas acadêmicas, e uma das formas de participação da sociedade para buscar ações e políticas públicas que promovam o desenvolvimento das regiões do estado de forma menos desigual é por meio dos Conselhos Regionais de Desenvolvimento (COREDES).

O principal objetivo desta pesquisa foi a constituição de um índice de desenvolvimento rural (IDR) para o COREDE central/RS, composto por 19 municípios, por meio da análise multivariada de 15 variáveis pertencentes a quatro dimensões de desenvolvimento: social, econômica, demográfica e ambiental.

O fenômeno chamado desenvolvimento rural é composto por várias dimensões, sendo a análise fatorial utilizada com a finalidade de captar os fatores componentes dessas dimensões. Dentre os dois fatores de desenvolvimento rural identificados nesta pesquisa, estão: Pessoas e Ambiente, e, Pessoas e Economia.

Primeiramente, foi feita a análise de agrupamento para verificar as similaridade das variáveis e se chegou a dois grupos reduzidos a nove variáveis, as quais se utilizou para a análise fatorial e análise dos componentes principais. Estas resultaram em dois fatores maiores que 1, que explicam 87,30% da variância total dos dados.

Analisando o desenvolvimento rural do COREDE central/RS, o IDR demonstrou que existem municípios com desigualdade de desenvolvimento dentro do COREDE central/RS. Pode-se perceber o perfil da região como tendo grande dependência do setor agropecuário com baixa diversidade da base econômica e certa concentração fundiária. Os dois IDRs retratam essas características regionais, quando em meio a múltiplas variáveis pertencentes a quatro dimensões de desenvolvimento, chega-se ao destaque para a produção de soja (SOJ) e arroz (ARR) e elevado valor adicionado bruto (VAB) no setor agropecuário. Sabe-se que os demais setores, comércio, serviços e indústrias são fomentados pelo setor rural.

Neste sentido, percebe-se que os municípios de Júlio de Castilhos e Tupanciretã representam o IDR2 com a produção de soja e alto VAB, e os municípios de Agudo e Santa Maria representam o IDR1 com o tratamento do solo (*ECP*) e população rural ocupada no ambiente rural (*POC*).

A presente pesquisa ficou limitada ao período de estudo, pois o ano de 2010 foi o último ano que ocorreu o censo do IBGE, restringindo assim a coleta de dados, porém é possível dar continuidade ao estudo a partir do próximo censo.

Por fim, torna-se necessário salientar a importância de estudar o desenvolvimento rural e fazer relações com outras temáticas, áreas e setores.

E como sugestão para futuras pesquisas pode-se citar a importância de analisar o desenvolvimento rural com número maior de variáveis construindo o IDR para cada município, região ou país, considerando mais dimensões que permitam maior precisão na análise para assim nortear as políticas públicas criadas para promover o desenvolvimento regional, como também expandir o universo de análise.

REFERÊNCIAS

- ALONSO, J. A. F. O cenário regional gaúcho nos anos 90: convergência ou mais desigualdade? **Indicadores Econômicos FEE**, Porto Alegre, v. 31, n. 3, p. 97-118, 2003.
- ALVES, L. B. **Índice de Desenvolvimento Rural dos Municípios Goianos: Uma análise de seus fatores determinantes**. Revista de Economia, Anápolis, v. 8, n. 2, p. 100-119, 2012.
- ANDERSON, T.W. **An introduction to multivariate statistical analysis**. New York: Wiley, 6 ed. 2003. 374p.
- BASSAB, W. de O.; MIAZAKI, E. S.; ANDRADE, D. F. de. **Introdução à Análise de Agrupamentos**. São Paulo: Associação Brasileira de Estatística (ABE), 1990.
- Campo e lavoura. **Gaúcha ZH**. Porto Alegre, 10 ago 2018. p. 30-49.
- CARVALHO, Marcia Cecília. **Construindo o Saber – Metodologia Científica: fundamentos e técnicas**. 19. ed. Campinas: Papyrus, 2008.
- CONTERATO, M. A.; SCHNEIDER, S. **Conformações regionais do desenvolvimento rural e da agricultura familiar: desigualdade e diversidade da geografia econômica do Rio Grande do Sul**. Revista Extensão Rural, Santa Maria, ano XVII, n. 19, p. 5-47, 2010.
- CONTERATO, M. A.; SCHNEIDER, S.; WAQUIL, P. D. **Desenvolvimento rural no Estado do Rio Grande do Sul: uma análise multidimensional de suas desigualdades regionais**. REDES, Santa Cruz do Sul, v. 12, n. 2, p. 163-195, 2007.
- CONTERATO, M. A.; SCHNEIDER, S.; WAQUIL, P. D. **Desigualdades regionais de desenvolvimento rural do Rio Grande do Sul: uma proposta de análise multidimensional a partir de três microrregiões**. Ensaio FEE, Porto Alegre, v. 30, número especial, p. 615-654, 2009.
- FERREIRA, D.F. **Estatística Multivariada**. Lavras: UFLA, 2011.
- FIGUEIREDO FILHO, D. B.; SILVA JUNIOR, J. A. **Visão além do alcance: uma introdução à análise fatorial**. Opin. Publica, Campinas, v.16, n.1, p. 160-185, jun. 2010. Disponível em http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-62762010000100007&lng=pt&nrm=iso. Acesso em outubro 2019. <https://doi.org/10.1590/S0104-62762010000100007>.
- FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA. FEE – Fundação de Economia e Estatística. Porto Alegre, 2003. Disponível em <www.fee.rs.gov.br>. Acesso em julho de 2018, outubro de 2018 e dezembro de 2018.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 5 ed. São Paulo: Atlas, 2010.

HAIR, J. F. et al. **Análise Multivariada de Dados**. 6. ed. Dados eletrônicos. Porto Alegre: Bookman, 2009.

HONGYU, K., SANDANIELO, V. L. M., JUNIOR, G. J. O. **Análise de Componentes Principais**: resumo teórico, aplicação e interpretação. E&S - Engineering and Science, v. 1, n. 5, 2016.

HONGYU, K. **Comparação do GGE- biplot ponderado e AMMI-ponderado com outros modelos de interação genótipo x ambiente**. 2015. 155p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Censo Demográfico de 2010. Disponível em: <<http://www.sidra.ibge.gov.br/cd/defaultcd2010.asp?o=4&i=P>>. Acesso em: maio de 2018.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 1998. 816p.

KAGEYAMA, A. **Desenvolvimento Rural**: conceito e medida. Cadernos de Ciência & Tecnologia, Brasília, DF, v. 21, n. 3, p. 379-408, 2004.

KAGEYAMA, A. **Desenvolvimento Rural**: conceitos e aplicações ao caso brasileiro. Porto Alegre: Editora da UFRGS, 2008.

LOESCH, C.; HOELTGEBAUM, M. **Métodos estatísticos multivariados**. São Paulo: Saraiva, 2012.

MALHOTRA, N. K. **Pesquisa de marketing**: uma orientação aplicada. 6. Ed – Dados eletrônicos. Porto Alegre: Bookman, 2012.

MANLY, B. F. J. **Multivariate statistical methods**. New York, Chapman and Hall, 1986. 159 p.

MELO, C. O. de; PARRÉ, J. L. **Índice de desenvolvimento rural dos municípios paranaenses**: determinantes e hierarquização. Revista de Economia e Sociologia Rural, Rio de Janeiro, v. 45, n. 2, p. 329-365, 2007.

MINGOTI, S. A. **Análise de dados através de métodos de Estatística Multivariada** — uma abordagem aplicada. Belo Horizonte: UFMG, 2007.

PINTO, N. G. M.; CORONEL, D. A. **Desenvolvimento rural no Rio Grande do Sul**: uma análise das mesorregiões entre 2000 e 2010. Ensaios FEE, Porto Alegre, v. 36, n. 4, p. 893-920, 2016.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico [recurso eletrônico]: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

RAMPAZZO, L. **Metodologia científica**: para alunos dos cursos de graduação e pós-graduação. São Paulo: Loyola, 2005.

SOUZA, A. M.; VICINI, L. **Análise multivariada: teoria e prática**. Santa Maria: UFSM/CCNE, 2005.

SECRETARIA DE PLANEJAMENTO, ORÇAMENTO E GESTÃO – **Atlas socioeconômico do Rio Grande do Sul**. <https://atlassocioeconomico.rs.gov.br/conselhos-regionais-de-desenvolvimento-coredes>. Acesso em 10/01/2020.

STEGE, A. L.; PARRE, J. L. **Desenvolvimento rural nas microrregiões do Brasil: um estudo multidimensional**. Teoria e Evidência Econômica, Passo Fundo, v. 17, p 160-193, 2011.

UFSM. **Estrutura e apresentação de monografias, dissertações e teses – MDT/Universidade Federal de Santa Maria**. Santa Maria: UFSM, 2015.

VASCONCELOS, V. V.; FREITAS, F. F. **Análise de agrupamentos (cluster analysis)**. Disponível em: https://pt.slideshare.net/vitor_vasconcelos/anlise-de-agrupamentos-clusters. Acessado em 18/01/2020.

VIALI, L. **Série estatística multivariada: texto I**. Porto Alegre: PUC-RS, 2015.

VICINI, L. et al. **Técnicas multivariadas exploratórias: teorias e aplicações no software estatística**. Santa Maria: UFSM, 2018.

VIEIRA, P. F. e MAIMON, D. (Organizadores). **As ciências sociais e a questão Ambiental: Rumo à interdisciplinaridade**. APED E UFPA, 1993.

WARD, J. H. **Hierarchical grouping to optimize an objective function**. Journal of the American statistical association, USA, v. 58, p. 236-244. DOI: 10.1080/01621459.1963.10500845. Acessado em janeiro de 2020.