

Componentes Principais:

Uma Análise dos COREDEs com perfil agropecuário do Rio Grande do Sul com base no Censo Agropecuário de 2017

Rafael Pentiado Poerschke

Prof. Dr. João R. Lazzarin¹

¹Centro de Ciências Naturais e Exatas (CCNE)
Universidade Federal de Santa Maria (UFSM)

15 de dezembro de 2023

Roteiro

Referências

Introdução

Metodologia

Desenvolvimento

Referências

▶ Referências Principais

- ▶ JOLLIFFE, Ian T. (1990). Principal component analysis: a beginner's guide—I. Introduction and application. *Weather*, v. 45, n. 10, p. 375-382.
- ▶ JOHNSON, R. A., WICHERN, D. W. and others. (2002) *Applied multivariate statistical analysis*, Prentice hall Upper Saddle River, NJ.

▶ Referências Complementares

- ▶ SCHNEIDER, S., WAQUIL, F. (2001). Caracterização socioeconômica dos municípios gaúchos e desigualdades regionais *Revista de Economia e Sociologia Rural*, 3(39), 117–142.
- ▶ STRANG, G. (2019). *Linear algebra and learning from data*, 1ed. SIAM.
- ▶ Instituto Brasileiro de Geografia e Estatística - IBGE. (2017) *Censo Agropecuário 2017*. Rio de Janeiro, RJ: IBGE, 2017. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhesid=73096>. Acesso em: 23 julho 2023.

Roteiro

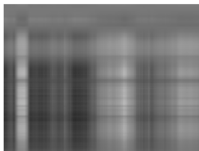
Referências

Introdução

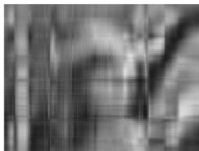
Metodologia

Desenvolvimento

Componentes Principais: Ideia principal/motivação



(a) 1 principal component



(b) 5 principal component



(c) 9 principal component



(d) 13 principal component



(e) 17 principal component



(f) 21 principal component

Questões centrais da pesquisa

- ▶ **COREDEs:** Conselhos Regionais de Desenvolvimento;
- ▶ **Decomposição em Valores Singulares:** Aplicação na Análise de Componentes Principais;
- ▶ **Componentes Principais:** Estimar para dados Sócio-Econômicos.

Questões centrais da pesquisa

- ▶ **COREDEs:** Conselhos Regionais de Desenvolvimento;
- ▶ **Decomposição em Valores Singulares:** Aplicação na Análise de Componentes Principais;
- ▶ **Componentes Principais:** Estimar para dados Sócio-Econômicos.

Questões centrais da pesquisa

- ▶ **COREDEs:** Conselhos Regionais de Desenvolvimento;
- ▶ **Decomposição em Valores Singulares:** Aplicação na Análise de Componentes Principais;
- ▶ **Componentes Principais:** Estimar para dados Sócio-Econômicos.

Problema de Pesquisa

O foco do presente estudo está em aplicar uma transformação linear que leve a **redução da dimensão** inicial de dados. Especificamente, intenta-se selecionar os **Componentes Principais**, com a utilização da linguagem estatística R a fim de investigar os COREDEs agropecuários gaúchos, considerando a existência de similaridade entre os municípios com base nos dados do Censo Agropecuário de 2017.

Problema de Pesquisa

Em um universo de 127 municípios, agregados em 8 COREDEs predominantemente **agropecuários**, questiona-se **o quão homogêneo será** esse grupo, isto é, **em que medida a agregação por contiguidade**, garantiria a homogeneidade dos COREDEs.

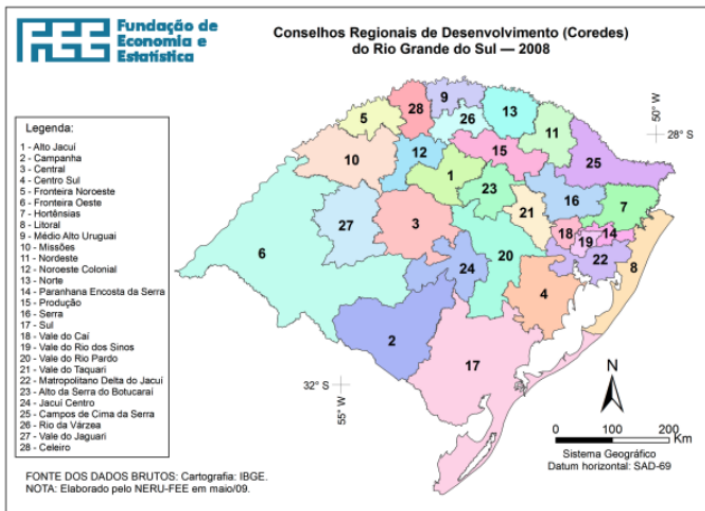
COREDEs: Origem

Definição: Conselhos Regionais de Desenvolvimento

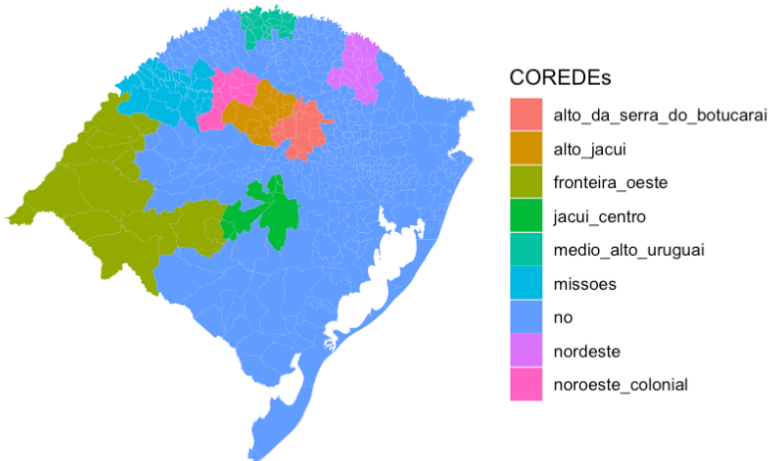
Os Conselhos Regionais de Desenvolvimento – COREDEs, criados oficialmente pela Lei 10.283 de 17 de outubro de 1994, são um fórum de discussão para a promoção de políticas e ações que visam o desenvolvimento regional.

Temos **28** COREDEs e **497** municípios no estado - começou com 21.

COREDEs: Localização



COREDEs Agropecuários



Hipótese

A agregação de um grupo de municípios no estado do RS por **contiguidade** - na forma dos COREDEs - não é suficiente para garantir a homogeneidade entre os municípios que fazem parte dos COREDEs agropecuários.

Objetivos

O **objetivo geral** deste trabalho é implementar a Análise de Componentes Principais para verificar a coesão entre os municípios partícipes dos COREDEs predominantes agrícolas do Rio Grande do Sul.

Objetivos Específicos

- ▶ Revisar a técnica de decomposição de **formas quadráticas** em sua representação matricial;
- ▶ Esquadrinhar a **Decomposição em Valores Singulares (DVS)** e sua aplicação em **Componentes Principais**;
- ▶ **Verificar empiricamente** a disposição dos municípios, inicialmente agrupados por adjacência, a fim de inferir sobre a coesão utilizando componentes principais;
- ▶ Estabelecer uma **nova dispersão**, no formato de grupos, desses municípios segundo os dados do Censo Agropecuário 2017 do IBGE selecionando os Componentes Principais de maior correlação com as variáveis utilizadas.

Roteiro

Referências

Introdução

Metodologia

Desenvolvimento

Tipologia da pesquisa

A condução deste trabalho baseia-se em uma abordagem quantitativa e matemática, buscando explorar a estrutura de similaridade entre os municípios dos COREDEs agrícolas do Rio Grande do Sul.

Roteiro

Referências

Introdução

Metodologia

Desenvolvimento

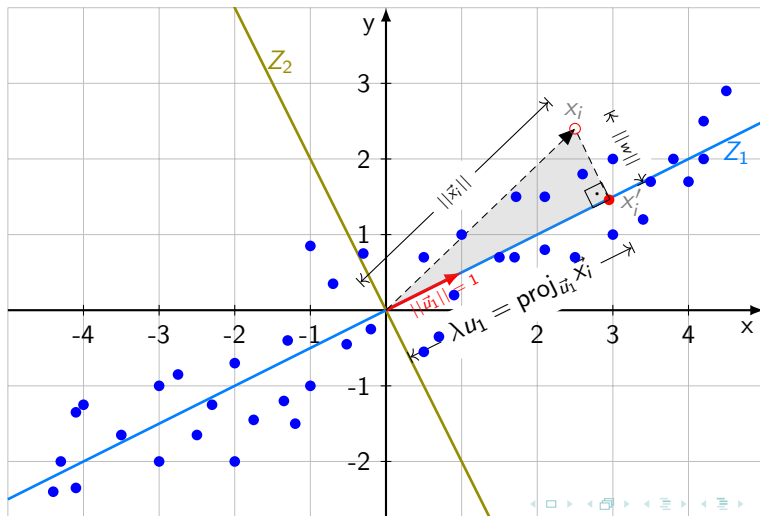
Componentes Principais

A Análise de Componentes Principais é um problema no qual busca-se estimar um subespaço de **dimensão inferior** $m \leq p$ de um conjunto de pontos em um espaço de dimensão maior \mathbb{R}^p dispostos em uma matriz $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ formada por p variáveis aleatórias correlacionadas entre si.

Componentes Principais

Esse problema pode ser modelado como uma questão **estatística** ou **geométrica**. Existe uma terceira abordagem, no qual ACP é vista como um problema de **aproximação** de uma matriz de **menor posto** em relação original.

O Problema da ACP



Componentes Principais

Teorema: Componentes Principais de Variáveis Aleatórias¹

Assuma que $\text{posto}(\mathbf{S}_X) \geq d$. Então os primeiros d componentes principais de uma variável aleatória multivariada \mathbf{X} , denotado por w_i para $i = 1, 2, \dots, d$, são dados por

$$w_i = u_i^T \mathbf{X},$$

onde $\{u_i\}_{i=1}^d$ são os d autovetores de $\mathbf{S}_X = \mathbb{E}[\mathbf{X}\mathbf{X}^T]$ associada com os d maiores autovalores $\{\lambda_i\}_{i=1}^d$. Além disso, $\lambda_i = \text{Var}(w_i)$ para $i = 1, 2, \dots, d$.

¹A demonstração do teorema pode ser consultada em Jolliffe (1990). 

Em resumo:

A transformação linear do vetor $\mathbf{X} = [X_1, X_2, \dots, X_p]$ de variáveis correlacionadas, que possui matriz de variâncias-covariâncias, será transformado em novas variáveis não-correlacionadas Y_1, Y_2, \dots, Y_p . As coordenadas dessas novas variáveis são descritas pelos vetores característicos \vec{e}_j de $\mathbf{W}_{(p \times p)}$ usados na seguinte transformação:

$$\underbrace{\mathbf{Y}}_{(p \times p)} = \underbrace{\mathbf{W}^t}_{(p \times p)} \underbrace{\mathbf{X}}_{(p \times p)} ; \quad (1)$$

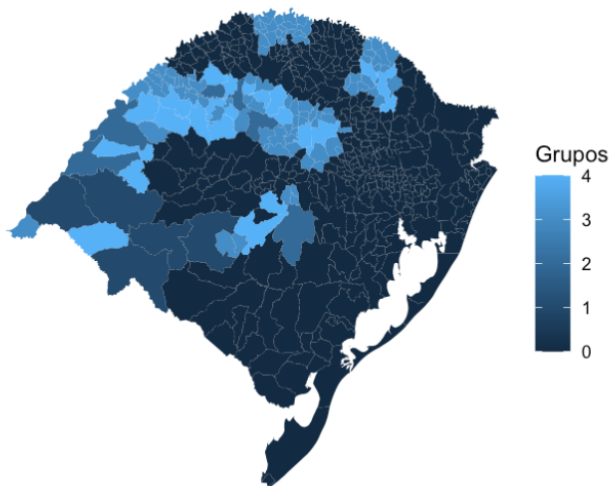
Em resumo:

Sendo **Y** a matriz de componentes principais, **W** possui os autovetores e **R** é a matriz de correlações das variáveis em **X**. Finalmente, podemos obter **F**, que é a nova matriz de escores das observações. Com isso, o problema se resume em

$$\underbrace{\mathbf{Y}}_{(m \times p)} = \underbrace{\mathbf{W}^t}_{(m \times p)} \underbrace{\mathbf{R}}_{(p \times p)} ; \quad (2)$$

$$\underbrace{\mathbf{F}}_{(n \times m)} = \underbrace{\mathbf{X}}_{(n \times p)} \underbrace{\mathbf{W}}_{(p \times m)} . \quad (3)$$

Grupos - COREDEs Agropecuários



Componentes Principais:

Uma Análise dos COREDEs com perfil agropecuário do Rio Grande do Sul com base no Censo Agropecuário de 2017

Rafael Pentiado Poerschke

Prof. Dr. João R. Lazzarin ¹

¹Centro de Ciências Naturais e Exatas (CCNE)
Universidade Federal de Santa Maria (UFSM)

15 de dezembro de 2023