# Análise de Componentes Principais (ACP):

## uma breve introdução

Rafael Pentiado Poerschke

Econometria II [1]

[1]Centro de Ciências Sociais e Humanas (CCSH)
Universidade Federal de Santa Maria (UFSM)

23 de março de 2024

# Roteiro

# 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.
  - FERREIRA, Daniel Furtado. (2008). **Estatística multivariada**. Lavras: Editora Ufla.
  - BIBBY, Ian L.; MARDIA, K. V.; KENT, John. (1993). **Multivariate shape analysis**. Sankhyā: The Indian Journal of Statistics, Series A, p. 460-480.


- Referências Complementares


  - ELDÉN, Lars. (2019). **Matrix methods in data mining and pattern recognition**. Society for Industrial and Applied Mathematics.
  - STRANG, G. (2019). **Linear algebra and learning from data**, 1ed. SIAM.
  - OLVER, Peter ; SHAKIBAN, Shakiban. (2018) **Applied Linear Algebra**. Springer

# Roteiro

## Componentes Principais: Ideia principal/motivação

**Dados em elevada dimensão:**

- Histórico das 400 empresas na B3
  - Categorias?
- Imagens: 9x13 cm $\Longrightarrow$ 1050x1500 pixels
  - Matriz menor?
- Histórico de buscas na Internet
  - *Ranqueamento* de páginas.

## Componentes Principais: Motivação

**Dados em elevada dimensão:**

- Histórico das 400 Empresas na B3
  - Categorias?
- Imagens: 9x13 cm $\Longrightarrow$ 1050x1500 pixels
  - Matriz menor? Sistema RGB.
- Histórico de buscas na Internet
  - *Ranqueamento* de páginas.

## Componentes Principais: Motivação

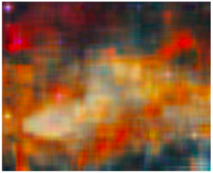
**Dados em elevada dimensão:**

- Histórico das 400 Empresas na B3
  - Categorias?
- Imagens: 9x13 cm $\Longrightarrow$ 1050x1500 pixels
  - Matriz menor? Sistema RGB.
- Histórico de buscas na Internet
  - *Ranqueamento* de páginas. Espaço Coluna.

# Componentes Principais: Motivação (780 x 960 pixels)

(a) original image

(b) $n = 1$. This is merely $\sigma_1 u_1 v_1^T$, the first term in the sum of 780 dyads $\sigma_i u_i v_i^T$. Thus, it bares very little resemblance to the original image.



(c) $n = 10$; $\sigma_1 u_1 v_1^T + \cdots + \sigma_{10} u_{10} v_{10}^T$
With only $\frac{10}{780}$ dyads, or 1.28% of all the data, we can already see the shapes in the original image starting to come together.

(d) $n = 100$; $\sigma_1 u_1 v_1^T + \cdots + \sigma_{100} u_{100} v_{100}^T$
Notice that with only $\frac{100}{780}$, or 13%, of the information contained in the original image we have an approximation that is close to being indistinguishable from the original.

## Questões centrais

- **Análise de Componentes Principais**: O que é? De onde vem? Como funciona?
- **Decomposição de Matrizes**: Como o software executa?
- **Economia do Desenvolvimento**: Aplicação da ACP.

## Questões centrais

- **Análise de Componentes Principais**: O que é? De onde vem? Como funciona?
- **Decomposição de Matrizes**: Como o software executa?
- **Economia do Desenvolvimento**: Aplicação da ACP.

# Questões centrais

- **Análise de Componentes Principais**: O que é? De onde vem? Como funciona?
- **Decomposição de Matrizes**: Como o software executa?
- **Economia do Desenvolvimento**: Aplicação da ACP.

# Análise de Componentes Principais

- **O que é** (para a Matemática): é uma transformação linear;
  - **Ingredientes**: ortogonalidade e decomposição em autovalores.
- **De onde vem**: Pearson (1901) e Hotelling (1933);
- **Para que serve**: possibilita a redução da dimensão inicial de dados.
- **Como faz**: APC é um modelo linear.

# Análise de Componentes Principais

- **O que é** (para a Matemática): é uma transformação linear;
  - **Ingredientes**: ortogonalidade e decomposição em autovalores.
- **De onde vem**: Pearson (1901) e Hotelling (1933);
- **Para que serve**: possibilita a redução da dimensão inicial de dados.
- **Como faz**: APC é um modelo linear.

# Análise de Componentes Principais

- **O que é** (para a Matemática): é uma transformação linear;
  - **Ingredientes**: ortogonalidade e decomposição em autovalores.
- **De onde vem**: Pearson (1901) e Hotelling (1933);
- **Para que serve**: possibilita a redução da dimensão inicial de dados.
- **Como faz**: APC é um modelo linear.

## Análise de Componentes Principais

- **O que é** (para a Matemática): é uma transformação linear;
  - **Ingredientes**: ortogonalidade e decomposição em autovalores.
- **De onde vem**: Pearson (1901) e Hotelling (1933);
- **Para que serve**: possibilita a redução da dimensão inicial de dados.
- **Como faz**: APC é um modelo linear.

# Análise de Componentes Principais

- **O que é** (para a Matemática): é uma transformação linear;
  - **Ingredientes**: ortogonalidade e decomposição em autovalores.
- **De onde vem**: Pearson (1901) e Hotelling (1933);
- **Para que serve**: possibilita a redução da dimensão inicial de dados.
- **Como faz**: APC é um modelo linear.

# Roteiro

# Problemas em Economia: Problema 1

Kageyama, Angela and Leone, Eugenia Troncoso. **Uma tipologia dos municípios paulistas com base em indicadores sociodemográficos**. Campinas: UNICAMP/IE, 1999.

## Problemas em Economia: Problema 1

- Kageyama e Leone (1999) buscaram construir uma **tipologia de economias regionais** em um conjunto de municípios (características sociais e econômicas).
- Para as autoras, a maioria dos estudos à época consistiam de **análises em profundidade** de localidades específicas:
  - **Objetivo**: geração de unidades territoriais maiores e mais homogêneas para o estudo das famílias agrícolas.

# Problemas em Economia: Problema 1

- Os resultados mostraram **cinco regiões** relativamente homogêneas no estado de São Paulo:
  - rural *muito* pobre;
  - rural *pobre*; *intermediária*;
  - *urbano* em expansão; e
  - *urbano* denso.

- Essas **tipificações** foram descritas em termos de:
  - renda,
  - população,
  - produção agrícola locais.

# Problemas em Economia: Problema 2

Schneider, Sérgio; Waquil, Paulo D. Caracterização socioeconômica dos municípios gaúchos e desigualdades regionais. **Revista de Economia e Sociologia Rural**. n.3, v.39. 2001. pp. 117-142.

- Os autores classificaram o Rio Grande do Sul em **cinco grupos** homogêneos de municípios, sendo que dois deles tinham em comum a pobreza rual e a degradação dos recursos naturais.

- **Conclusão**: descartaram a ideia de que o estado podia ser dividido em duas partes, isto é, entre uma *metade sul* **mais atrasada**, e o *norte* **desenvolvido**.

**ver também**: Freitas et al., Analisando a modernização da agropecuária gaúcha: uma aplicação de análise fatorial e cluster.

# Problemas em Economia: Problema 3

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

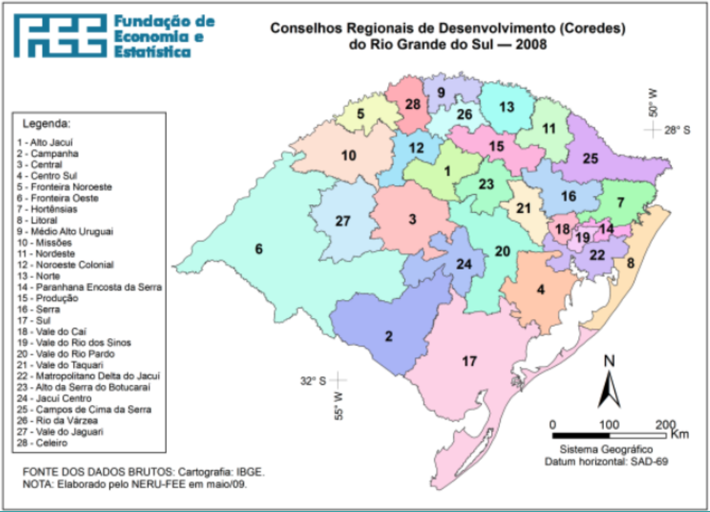
# COREDEs: Origem

## Definição: Conselhos Regionais de Desenvolvimento

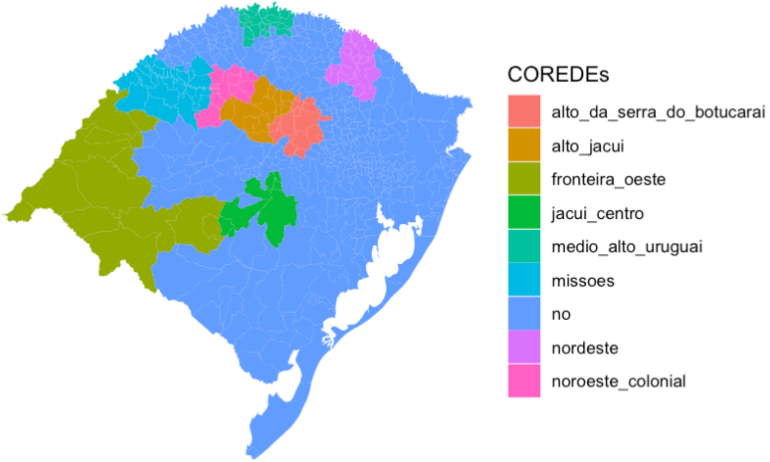
Os Conselhos Regionais de Desenvolvimento – COREDEs, criados oficialmente pela Lei 10.283 de 17 de outubro de 1994, são um fórum de discussão para a promoção de políticas e ações que visam o desenvolvimento regional.

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

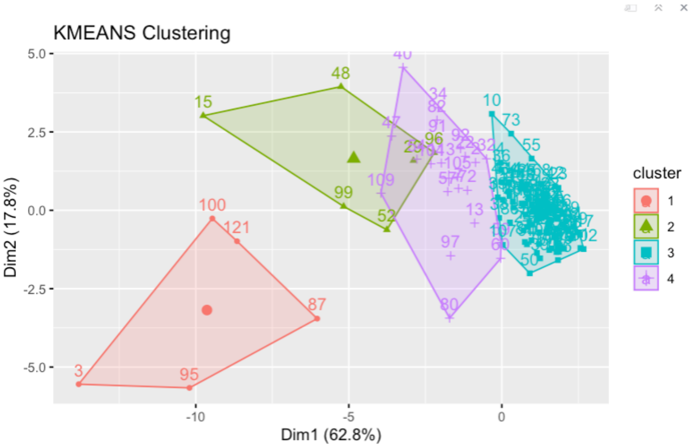
# COREDEs: Localização



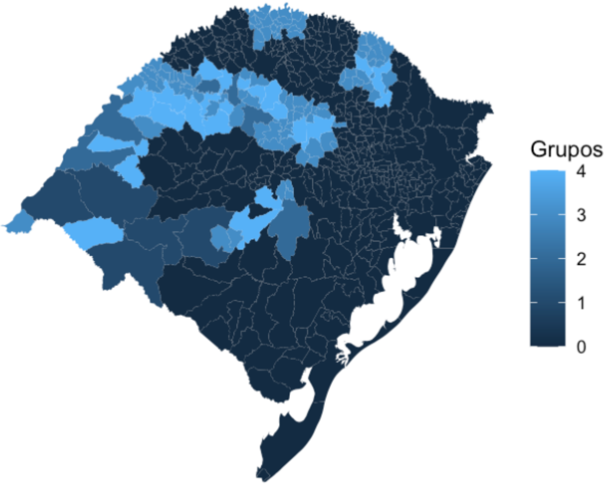
# COREDEs Agropecuários



# Grupos - COREDEs Agropecuários



# Grupos - COREDEs Agropecuários



# Roteiro

## 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 $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_p]^T$ formada por $p$ variáveis aleatórias **correlacionadas entre si**.

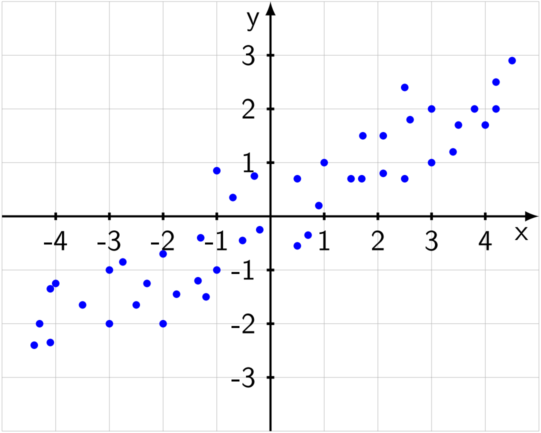
Em termos de transformação linear, significa

$$T : \mathbb{R}^p \to \mathbb{R}^m$$

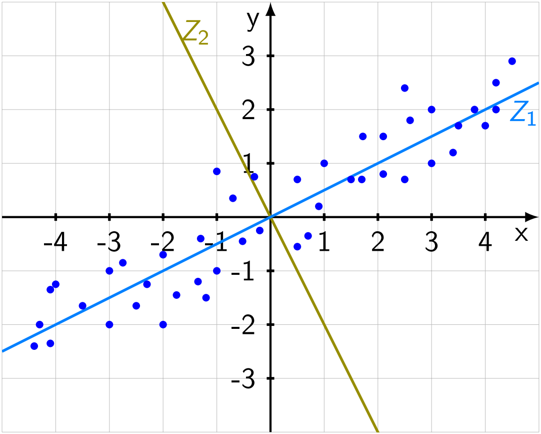
## Componentes Principais

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.

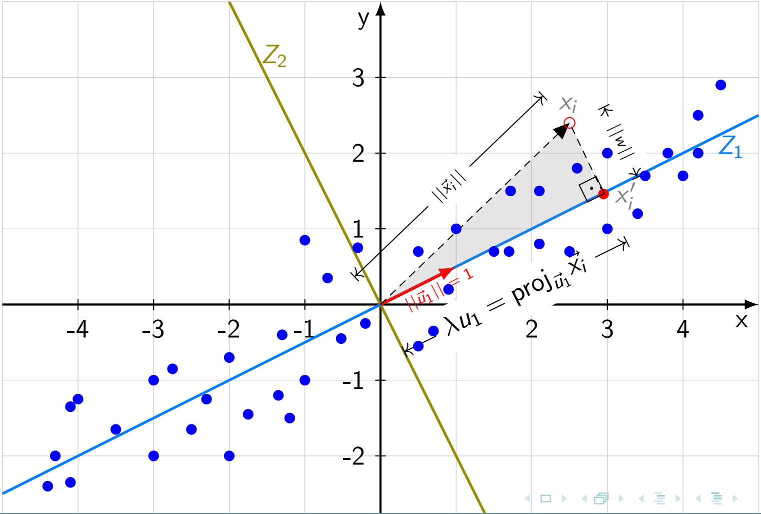
# O Problema da ACP: abordagem geométrica



# O Problema da ACP: abordagem geométrica



# O Problema da ACP: abordagem geométrica



# Componentes Principais: abordagem geométrica

Dada uma variável aleatória multivariada de média zero $\mathbf{X} \in \mathbb{R}^p$ e um inteiro $m < p$, os $m$ "componentes principais"de $\mathbf{X}$, $\mathbf{w} \in \mathbb{R}^m$, são definidos como os $m$ componentes lineares não correlacionados de $\mathbf{X}$,

$$\mathbf{w}_i = \mathbf{u}_i^\top \mathbf{X}, \quad i \in \{1, 2, \ldots, m\},$$

onde $\mathbf{u}_i \in \mathbb{R}^p$, para $i = 1, 2, \ldots, m$.
Tal que a variância de $w_i$ é maximizada sujeita a

$$\mathbf{u}_i^\top \mathbf{u}_i = 1 \text{ e } \mathrm{Var}(\mathbf{w}_1) \geq \mathrm{Var}(\mathbf{w}_2) \geq \cdots \geq \mathrm{Var}(\mathbf{w}_m) \geq 0.$$

Por exemplo, para encontrar o primeiro componente principal $\mathbf{w}_1$, procuramos um vetor $\mathbf{u}_1 \in \mathbb{R}^p$ tal que

$$\mathbf{u}_1 = \arg \max_{\mathbf{u}_1 \in \mathbb{R}^p} \mathrm{Var}(\mathbf{u}_1^\top \mathbf{x}) \quad \text{s.a.} \quad \mathbf{u}_1^\top \mathbf{u}_1 = 1.$$

O teorema a seguir mostra que os componentes principais de $\mathbf{x}$ podem ser computados a partir dos autovetores de sua matriz de covariância $\mathbf{S}_x = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$.

## Componentes Principais: abordagem geométrica

**Teorema**: Componentes Principais de Variáveis Aleatórias[1]

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

$$\mathbf{w}_i = \mathbf{u}_i^\top \mathbf{X},$$

onde $\{\mathbf{u}_i\}_{i=1}^m$ são os $m$ autovetores de $\mathbf{S}_X = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ associada com os $m$ maiores autovalores $\{\lambda_i\}_{i=1}^m$. Além disso, $\lambda_i = \mathrm{Var}(\mathbf{w}_i)$ para $i = 1, 2, \ldots, m$.

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

## Componentes Principais: abordagem geométrica

**Demonstração:** Seja $\mathbf{X}$ um vetor de $p$-variáveis aleatórias correlacionadas, com $\mathbf{X} \in \mathbb{R}^p$ e $\mathbf{u}_i$ é um vetor de $p$ constantes $u_{i1}, u_{i2}, \ldots, u_{ip}$. Como objetivo, queremos preservar o máximo de informação possível em uma dimensão menor que a original, sendo a variância de $\mathbf{X}$ computada em $m$ Componentes Principais em que $m \leq p$, tal que $\mathbf{w}$ representa os $m$ componentes não correlacionados de $\mathbf{X}$ e $\mathbf{w} \in \mathbb{R}^m$. Assim,

$$\mathbf{w}_i = \mathbf{u}_i^{\mathsf{T}} \mathbf{X} \in \mathbb{R}, \quad u_i \in \mathbb{R}^p, \quad i = 1, 2, \ldots, m$$

tal que a variância de $\mathbf{w}_i$ é maximizada sujeita a

$$\mathbf{u}_i^{\mathsf{T}} \mathbf{u}_i = 1 \quad \text{e} \quad \text{Var}(\mathbf{w}_1) \geq \text{Var}(\mathbf{w}_2) \geq \ldots \text{Var}(\mathbf{w}_m) > 0.$$

# Componentes Principais: abordagem geométrica

Primeiro,vamos assumir que $\mathbf{S}_X$ não possui autovalores repetidos. Por ser uma matriz com entradas reais e simétrica, esses autovalores são reais e seus autovetores associados formam um base em $\mathbb{R}^p$. Mais que isso, os autovalores são únicos (positivos), e os autovetores correspondentes são ortogonais entre si.

Observe que para cada $\mathbf{u} \in \mathbb{R}^p$ temos que

$$\mathrm{Var}(\mathbf{u}_1^\mathsf{T}\mathbf{X}) = \tfrac{1}{n}\sum_{i=1}^{n}(u_1^\mathsf{T}x_i - u_1^\mathsf{T}\overline{x})^2.$$

# Componentes Principais: abordagem geométrica

Uma vez que os dados foram normalizados ao redor da média, a expressão se reduz a

$$\mathrm{Var}(\mathbf{u}_1^\mathsf{T}\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{u}_1^\mathsf{T}\mathbf{X}_i)^2$$

$$\mathrm{Var}(\mathbf{u}^\mathsf{T}\mathbf{X}) = \mathbb{E}[(\mathbf{u}^\mathsf{T}\mathbf{X})^2] = \mathbb{E}[\mathbf{u}^\mathsf{T}\frac{\mathbf{X}\mathbf{X}^\mathsf{T}}{n}\mathbf{u}] = \mathbf{u}^\mathsf{T}\mathbf{S}_X\mathbf{u}.$$

## Componentes Principais: abordagem geométrica

Vamos assumir que o queremos projetar os dados em um espaço unidimensional. O componente principal correspondente a esse eixo será combinação linear das variáveis originais expressa por $\mathbf{u}_i^T \mathbf{X}$, isto é:

$$\mathbf{w}_i = \mathbf{u}_i^T \mathbf{X} = u_{i1}\mathbf{X}_1 + u_{i2}\mathbf{X}_2 + \cdots + u_{ip}\mathbf{X}_p \quad i = 1, 2, \ldots, p;$$

isto é

$$\begin{aligned} w_{1i} &= u_{1i}\mathbf{X}_{11} + u_{2i}\mathbf{X}_{12} + \cdots + u_{pi}\mathbf{X}_{1p} \\ w_{2i} &= u_{1i}\mathbf{X}_{21} + u_{2i}\mathbf{X}_{22} + \cdots + u_{pi}\mathbf{X}_{2p} \\ &\vdots \\ w_{ni} &= u_{1i}\mathbf{X}_{n1} + u_{2i}\mathbf{X}_{n2} + \cdots + u_{pi}\mathbf{X}_{np} \,. \end{aligned}$$

## Componentes Principais: abordagem geométrica

Após, procuramos por um segundo componente, $\mathbf{u}_2^\top \mathbf{X}$, que não correlacionado com $\mathbf{u}_1^\top \mathbf{X}$ e tenha máxima variância, e assim sucessivamente até encontrarmos o $k$-ésimo estágio que possua máxima variância e ausência de correlação com $\mathbf{u}_1^\top \mathbf{X}, \mathbf{u}_2^\top \mathbf{X}, \ldots, \mathbf{u}_{k-1}^\top \mathbf{X}$.

## Componentes Principais: abordagem geométrica

Para derivarmos os Componentes Principais, consideremos um primeiro $\mathbf{u}_1^\top \mathbf{X}$; o vetor $\mathbf{u}_1$ maximiza $\mathrm{Var}(\mathbf{u}_1^\top \mathbf{X}) = \mathbf{u}_1^\top \mathbf{S}_X \mathbf{u}_1$. Como foi supra-exposto, $\mathbf{u}_i^\top \mathbf{u}_i = 1$ é assumido como restrição, e para solução fazemos uso do Lagrangiano.

Nesse sentido, tempos o problema de otimização condicionada dado por

$$\begin{cases} \max\limits_{\mathbf{u}_1 \in \mathbb{R}^p} & (\mathbf{u}_1^\top \mathbf{S}_X \mathbf{u}_1) \\ \text{sujeito a:} & ||\mathbf{u}_1|| = \mathbf{u}_1^\top \mathbf{u}_1 = 1. \end{cases} \tag{1}$$

Assim, para maximizar $\mathbf{u}_1^\top \mathbf{S}_X \mathbf{u}_1$ sujeito a $\mathbf{u}_k^\top \mathbf{u}_k = 1$, significa maximizar

$$\mathcal{L}(\mathbf{S}, \mathbf{u}_1, \lambda_1) = \mathbf{u}_1^\top \mathbf{S}_X \mathbf{u}_1 - \lambda_1(\mathbf{u}_1^\top \mathbf{u}_1 - 1),$$

onde $\lambda_1 \in \mathbb{R}$ representa o multiplicador de Lagrange.

# Teorema de Lagrange

**Teorema**: Método dos Multiplicadores de Lagrange[2]

Suponha que $f : U \subset \mathbb{R}^n \to \mathbb{R}$ e $g : U \subset \mathbb{R}^n \to \mathbb{R}$ são funções reais de classe $C^1$ no aberto $U \in \mathbb{R}^n$. Seja $\vec{x}_0 \in U$ e $g(\vec{x}_0) = c$, e seja $S$ o nível de $g$ ao valor $c$. Assuma que $\nabla g(\vec{x}_0) \neq \vec{0}$.
Se $f|S$ ($f$ restrita a $S$), possui um local máximo ou mínimo em $S$ no ponto $\vec{x}_0$, então existe um número real $\lambda$ que

$$\boxed{\nabla f(\vec{x}_0) = \lambda \nabla g(\vec{x}_0)}. \tag{2}$$

[2] Uma demonstração pode ser encontrada em MARSDEN, Jerrold E.; TROMBA, Anthony. Vector calculus. Macmillan, 2003.

# Teorema de Lagrange: derivando matrizes

$$g(\vec{x}_0) : \mathbf{u}^\mathsf{T}\mathbf{u} = 1$$

$$\begin{aligned}\mathbf{u}^\mathsf{T}\mathbf{u} &= \begin{bmatrix} u_1, & u_2, & \dots, & u_p \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{bmatrix} \\ &= \begin{bmatrix} u_1^2 + u_2^2 + \cdots + u_p^2 \end{bmatrix}.\end{aligned}$$

# Teorema de Lagrange

Sendo assim, precisamos encontrar $\nabla g(\vec{x}_0)$, tal que:

$$\frac{\partial}{\partial \mathbf{u}}\left(\mathbf{u}^{\mathsf{T}}\mathbf{u}\right) = \begin{bmatrix} \frac{\partial}{\partial u_1}\left(\mathbf{u}^{\mathsf{T}}\mathbf{u}\right) \\ \frac{\partial}{\partial u_2}\left(\mathbf{u}^{\mathsf{T}}\mathbf{u}\right) \\ \vdots \\ \frac{\partial}{\partial u_p}\left(\mathbf{u}^{\mathsf{T}}\mathbf{u}\right) \end{bmatrix}$$

# Teorema de Lagrange

Efetuando a derivada, coordenada a coordenada, temos

$$\frac{\partial}{\partial u_1}\left(\mathbf{u}^\mathsf{T}\mathbf{u}\right) = 2u_1;$$
$$\frac{\partial}{\partial u_2}\left(\mathbf{u}^\mathsf{T}\mathbf{u}\right) = 2u_2;$$
$$\vdots$$
$$\frac{\partial}{\partial u_p}\left(\mathbf{u}^\mathsf{T}\mathbf{u}\right) = 2u_p.$$

# Teorema de Lagrange

Montando o gradiente de $g(\vec{x}_0)$ encontramos:

$$\nabla(\mathbf{u}^{\mathsf{T}}\mathbf{u}) = [2u_1,\, 2u_2,\, \ldots,\, 2u_p]$$

$$= 2[u_1\,,\, u_2\,,\ldots,\, u_p]$$

$$\boxed{\nabla(\mathbf{u}^{\mathsf{T}}\mathbf{u}) = 2\mathbf{u}}$$

# Teorema de Lagrange

Agora busca-se o $\nabla f(\vec{x}_0)$ derivamos coordenada a coordenada:

$$\mathbf{u}^{\mathrm{T}}\mathbf{S}\mathbf{u} = \begin{bmatrix} u_1, & u_2, & \dots, & u_p \end{bmatrix}_{1\times p} \begin{bmatrix} \sigma^2_{11} & \sigma^2_{12} & \dots & \sigma^2_{1p} \\ \sigma^2_{21} & \sigma^2_{12} & \dots & \sigma^2_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{p1} & \sigma^2_{p2} & \dots & \sigma^2_{pp} \end{bmatrix}_{p\times p} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{bmatrix}_{p\times 1}$$

## Teorema de Lagrange

Então, iremos derivar coordenada a coordenada, mas se $\mathbf{u}^\mathsf{T}\mathbf{S}\mathbf{u} = \mathbf{u}^\mathsf{T}\mathbf{W}$, com $\mathbf{W} = \mathbf{S}\mathbf{u}$, precisamos aplicar a regra do produto em (0a), ou seja:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{u}}(\mathbf{u}^\mathsf{T}\mathbf{W}) &= \mathbf{u}^\mathsf{T}\frac{\partial \mathbf{W}}{\partial \mathbf{u}} + \mathbf{W}^\mathsf{T}\frac{\partial \mathbf{u}}{\partial \mathbf{u}} \\
&= \mathbf{u}^\mathsf{T}\mathbf{S} + \mathbf{W}^\mathsf{T} \\
&= \mathbf{u}^\mathsf{T}\mathbf{S} + \mathbf{u}^\mathsf{T}\mathbf{S}^\mathsf{T} \\
&= \mathbf{u}^\mathsf{T}(\mathbf{S} + \mathbf{S}^\mathsf{T}).
\end{aligned}$$

## Teorema de Lagrange

Como $\mathbf{S}$ é matriz de variância-covariância, portanto ela é simétrica ($\mathbf{S} = \mathbf{S}^\mathsf{T}$), logo:

$$\frac{\partial}{\partial u_1}(\mathbf{u}^\mathsf{T}\mathbf{S}\mathbf{u}) = 2\sum_{j=1}^{p} \sigma_{1j}^2 u_j$$

$$\frac{\partial}{\partial u_2}(\mathbf{u}^\mathsf{T}\mathbf{S}\mathbf{u}) = 2\sum_{j=1}^{p} \sigma_{2j}^2 u_j$$

$$\vdots$$

$$\frac{\partial}{\partial u_p}(\mathbf{u}^\mathsf{T}\mathbf{S}\mathbf{u}) = 2\sum_{j=1}^{p} \sigma_{pj}^2 u_j$$

$$\boxed{\frac{\partial}{\partial \mathbf{u}}(\mathbf{u}^\mathsf{T}\mathbf{S}\mathbf{u}) = 2\mathbf{S}\mathbf{u}}$$

# Componentes Principais: abordagem geométrica

**UFA! Agora sim podemos seguir!**

$$\mathcal{L}(\mathbf{S}, \mathbf{u}_1, \lambda_1) = \mathbf{u}_1^T \mathbf{S}_X \mathbf{u}_1 - \lambda_1(\mathbf{u}_1^T \mathbf{u}_1 - 1),$$

Diferenciando com respeito a $\mathbf{u}_1$, temos

$$2\mathbf{S}_X \mathbf{u}_1 - 2\lambda_1 \mathbf{u}_1 = 0$$

$$\mathbf{S}_X u_1 = \lambda_1 u_1$$

ou seja,

$$(\mathbf{S}_X - \lambda_1 \mathbf{I}_p)\mathbf{u}_1 = 0,$$

tal que $\mathbf{I}_p$ é a matriz identidade de dimensão $p \times p$. Então, $\lambda_1$ é um autovalor de $\mathbf{S}_X$ e $\mathbf{u}_1$ o seu autovetor correspondente.

# Autovalor: o nosso lambda[3]

A matriz $\mathbf{A} \in \mathbf{M}_{n\times n}(\mathbb{R})$ possui um autovalor $\lambda \in \mathbb{C}$ se, e somente se, existe um vetor $\underbrace{\mathbf{e}}_{(n\times 1)}$ não nulo tal que, $\mathbf{Ae} = \lambda\mathbf{e}$ . Nesse caso, dizemos que $\lambda$ é um autovalor associado ao autovetor $\mathbf{e}$.
Portanto, o problema de encontrar autovalores de uma matriz se reduz na procura pelas raízes de um certo polinômio em $\lambda$, denominado polinômio característico. Mais explicitamente,

$$\begin{aligned} \mathbf{Ae} = \lambda\mathbf{e} &\iff \mathbf{Ae} - \lambda\mathbf{e} = 0 \\ &\iff (\mathbf{A} - \lambda\mathbf{I})\mathbf{e} = 0 \end{aligned} \tag{3}$$

[3] HOFFMAN, Kenneth; KUNZE, Ray. **Linear Algebra**, Engelewood Cliffs. 1961.

# Autovalor: o nosso lambda

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{e} \; = \; 0$$

como $\mathbf{e} \neq 0$ o sistema $[\mathbf{A} - \lambda \mathbf{I}]\,\mathbf{e} = 0$ possui solução não nula se, e somente, se

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0 \tag{4}$$

tal que: $\mathbf{A} = \; \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T \; + \; \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T \; + \; \cdots \; + \; \lambda_k \mathbf{e}_k \mathbf{e}_k^T$[4]

[4]Conhecida como Decomposição Espectral, que pode ser escrita também como: $\sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i^T \; = \; \mathbf{P \Lambda P}^{-1}$, tal que $\mathbf{\Lambda}$ é uma matriz diagonal de autovalores de $\mathbf{A}$. Por definição, uma matriz diagonal é uma matriz quadrada que possui todos os elementos acima e abaixo da diagonal principal nulos.

# Autovalor: o nosso lambda

**Definição:** Seja $V$ um espaço vetorial sobre o campo $F$ e $T$ um operador linear em $V$ tal que $T : V \to V$. Um **autovalor** de $\mathbf{T}$ é um escalar $\lambda$ em $F$ tal que existe um vetor **não-nulo**[5] $\mathbf{e} \in V$ com $\mathbf{Te} = \lambda\mathbf{e}$. Se $\lambda$ é um autovalor de $\mathbf{T}$, então

1. Qualquer $\mathbf{e}$ tal que $\mathbf{Te} = \lambda\mathbf{e}$ é chamado de **autovetor** de $\mathbf{T}$ associado ao autovalor $\lambda$;
2. A coleção de todos os $\mathbf{e}$ tais que $\mathbf{Te} = \lambda\mathbf{e}$ é chamada de **espaço característico associado** a $\lambda$.

---

[5] O vetor nulo não pode estar no espaço nulo de $\mathbf{T}$. **Espaço Nulo** $\mathcal{N}(\mathbf{T}) \in R^m$: é o conjunto de todas as soluções para o sistema $Ax = 0$. O núcleo de uma transformação linear ($Ker(\mathbf{T})$) é sinônimo, isto é, $Ker(\mathbf{T}) = \{\mathbf{e} \in V : \mathbf{Te} = 0\}$. Ambos se referem ao conjunto de vetores do domínio que são mapeados para o vetor nulo no contradomínio pela transformação linear.

# Autovalor: o nosso lambda

## Autovalores:

Os **autovalores** são frequentemente chamados: valores característico, raízes características, raízes latentes, valores próprios ou valores espectrais.

# Autovalor: o nosso lambda

Se $\mathbf{T}$ é qualquer operador linear e $\lambda$ é qualquer escalar, o conjunto de vetores $\mathbf{e}$ tal que $\mathbf{Te} = \lambda\mathbf{e}$ é um subespaço de $V$.

$$(\mathbf{T}-\lambda\mathbf{I}) \text{ é não injetiva } \iff \exists \mathbf{v}_1, \mathbf{v}_2 : \mathbf{v}_1 \neq \mathbf{v}_2 \text{ e } (\mathbf{T}-\lambda\mathbf{I})\mathbf{v}_1 = (\mathbf{T}-\lambda\mathbf{I})\mathbf{v}_2$$

$$\iff \exists \mathbf{x} : x \neq 0 \text{ e } (\mathbf{T}-\lambda\mathbf{I})\mathbf{x} = 0$$

$$\iff (\mathbf{T}-\lambda\mathbf{I}) \text{ é singular}$$

$$\iff \det(\mathbf{T}-\lambda\mathbf{I}) = 0$$

## Autovalor: o nosso lambda

Se $(\mathbf{T} - \lambda \mathbf{I})$ é não-singular, note que em

$$(\mathbf{T} - \lambda \mathbf{I})\mathbf{x} = 0, \quad \text{podemos multiplicar pela inversa}$$

$$(\mathbf{T} - \lambda \mathbf{I})^{-1}(\mathbf{T} - \lambda \mathbf{I})\mathbf{x} = (\mathbf{T} - \lambda \mathbf{I})^{-1}\, 0$$

$$\mathbf{I}\mathbf{x} = 0 \implies \mathbf{x} = 0$$

Exatamente o que não queremos.

## Autovalor: o nosso lambda

**Teorema** Seja $\mathbf{T}$ um operador linear em um espaço de dimensão finita $V$ e $\lambda$ um escalar. As seguintes afirmações são equivalentes.

(i) $\lambda$ é um valor característico de $\mathbf{T}$.

(ii) O operador $(\mathbf{T} - \lambda\mathbf{I})$ é singular (não invertível).

(iii) $\det(\mathbf{T} - \lambda\mathbf{I}) = 0$.

O critério do determinante (iii) é muito importante, pois nos diz onde procurar os valores característicos de $\mathbf{T}$. Como $\det(\mathbf{T} - \lambda\mathbf{I})$ é um polinômio de grau $n$ na variável $\lambda$, encontraremos os valores característicos como as raízes desse polinômio.

# Autovalor: o nosso lambda

Se $\mathcal{B}$ é qualquer base ordenada para $V$ e $\mathbf{A} = [T]_{\mathcal{B}}$, então $(\mathbf{T} - \lambda\mathbf{I})$ é inversível se e somente se a matriz $(\mathbf{A} - \lambda\mathbf{I})$ é inversível. De acordo com isso, fazemos a seguinte definição.

**Definição.** Se $\mathbf{A}$ é uma matriz $n \times n$ sobre o campo $F$, um valor característico de $\mathbf{A}$ em $F$ é um escalar $\lambda$ em $F$ tal que a matriz $(\mathbf{A} - \lambda\mathbf{I})$ é singular (não invertível).

Uma vez que $c$ é um valor característico de $\mathbf{A}$ se e somente se $\det(\mathbf{A} - c\mathbf{I}) = 0$, ou equivalentemente se e somente se $\det(c\mathbf{I} - \mathbf{A}) = 0$.

# Propriedades de autovetores de matrizes simétricas:

1. **Simetria:**
   - Em uma matriz simétrica $A$, os autovetores correspondentes a diferentes autovalores são ortogonais entre si.
2. **Diagonalização:**
   - Matrizes simétricas são diagonalizáveis, o que significa que podem ser decompostas na forma $A = PDP^{-1}$, onde $P$ é uma matriz formada pelos autovetores de $A$ e $D$ é uma matriz diagonal com os autovalores correspondentes.
3. **Autovalores Reais:**
   - Todos os autovalores de uma matriz simétrica são reais.
4. **Quantidade de Autovetores:**
   - Se uma matriz simétrica tem dimensão $n$, então ela possui exatamente $n$ autovetores linearmente independentes.

## Componentes Principais: abordagem geométrica

Para decidirmos qual dos $p$ autovetores produz $\mathbf{u}_1^\top$ de máxima variância, note que a quantidade a ser maximizada é

$$\mathbf{u}_1^\top \mathbf{S}_X \mathbf{u}_1 = \mathbf{u}_1^\top \lambda_1 \mathbf{u}_1 = \lambda_1 \mathbf{u}_1^\top \mathbf{u}_1 = \lambda_1,$$

assim, $\lambda_1$ deve ser o maior possível. Logo, $\mathbf{u}_1$ é o autovetor associado ao maior autovalor de $\mathbf{S}_X$, e
$\text{Var}(\mathbf{u}_1^\top \mathbf{S}_X \mathbf{u}_1) = \mathbf{u}_1^\top \lambda_1 \mathbf{u}_1 = \lambda_1$. Então, $\lambda_1 = \text{Var}(\mathbf{w}_1) > 0$.

## Componentes Principais: abordagem geométrica

Agora, para $k = 2$, o segundo componente $\mathbf{u}_2^\mathsf{T}\mathbf{X}$, maximiza $\mathbf{u}_2^\mathsf{T}\mathbf{S}\mathbf{u}_2$ que deve ser ausente de correlação com $\mathbf{u}_1^\mathsf{T}\mathbf{X}$, ou equivalente a a $\mathrm{cov}(\mathbf{u}_1^\mathsf{T}\mathbf{X}, \mathbf{u}_2^\mathsf{T}\mathbf{X}) = 0$, onde $\mathrm{cov}(x, y)$ denota a covariância entre as variáveis $x$ e $y$. Mas

$$\mathrm{cov}(\mathbf{u}_1^\mathsf{T}\mathbf{X}, \mathbf{u}_2^\mathsf{T}\mathbf{X}) = \mathbf{u}_1^\mathsf{T}\mathbf{S}_X\mathbf{u}_2 = \mathbf{u}_2^\mathsf{T}\mathbf{S}_X\mathbf{u}_1 = \mathbf{u}_2^\mathsf{T}\lambda_1\mathbf{u}_1^\mathsf{T} = \lambda_1\mathbf{u}_2^\mathsf{T}\mathbf{u}_1 = \lambda_1\mathbf{u}_1^\mathsf{T}\mathbf{u}_2 = 0$$

## Componentes Principais: abordagem geométrica

Assim, cada dessas equações

$$\mathbf{u}_1^{\mathsf{T}}\mathbf{S}_X\mathbf{u}_2 = 0$$
$$u_2^{\mathsf{T}}\mathbf{S}_X\mathbf{u}_1 = 0$$
$$\mathbf{u}_1^{\mathsf{T}}\mathbf{u}_2 = 0$$
$$\mathbf{u}_2^{\mathsf{T}}\mathbf{u}_1 = 0$$

pode ser usada para indicar ausência de correlação entre $\mathbf{u}_1^{\mathsf{T}}\mathbf{X}$ e $\mathbf{u}_2^{\mathsf{T}}\mathbf{X}$. Assumimos igualmente a hipótese de normalização entre $\mathbf{u}_1^{\mathsf{T}}\mathbf{X}$ e $\mathbf{u}_2^{\mathsf{T}}\mathbf{X}$, a quantidade que deve ser maximizada será

$$\mathcal{L}(\mathbf{S}, \mathbf{u}_1, \mathbf{u}_2, \lambda_2, \phi) = \mathbf{u}_2^{\mathsf{T}}\mathbf{S}_X\mathbf{u}_2 - \lambda_2(\mathbf{u}_2^{\mathsf{T}}\mathbf{u}_2 - 1) - \phi\mathbf{u}_2^{\mathsf{T}}\mathbf{u}_1,$$

onde $\lambda_2$ e $\phi$ são os multiplicadores de Lagrange.

## Componentes Principais: abordagem geométrica

Diferenciando a expressão em relação a $\mathbf{u}_2$, temos

$$\mathbf{S}_X\mathbf{u}_2 - \lambda_2\mathbf{u}_2 - \frac{\phi}{2}\mathbf{u}_1 = 0$$

e multiplicando a equação por $\mathbf{u}_1^\top$ temos

$$\begin{aligned} \mathbf{u}_1^\top\mathbf{S}_X\mathbf{u}_2 - \lambda_2\mathbf{u}_1^\top\mathbf{u}_2 - \phi\mathbf{u}_1^\top\mathbf{u}_1 &= 0 \\ 0 \; - \; 0 \; - \phi 1 \; &= \; 0, \end{aligned}$$

sendo que os dois primeiros termos são zero e $\mathbf{u}_1^\top\mathbf{u}_1 = 1$, temos $\phi = 0$.

## Componentes Principais: abordagem geométrica

Além disso, $\mathbf{S}_X\mathbf{u}_2 - \lambda_2\mathbf{u}_2 = 0$, ou equivalente a $(\mathbf{S}_X - \lambda_2\mathbf{I}_p)\mathbf{u}_2 = 0$, então $\lambda_2 = \mathrm{Var}(\mathbf{w}_2)$, e desde que os autovalores de $\mathbf{S}_X$ são distintos, $\lambda_2$ é o autovalor de $\mathbf{S}_X$, e $\mathbf{u}_2$ representa seu autovetor correspondente.

# Componentes Principais: abordagem geométrica

Esse processo pode ser repetido para $k = 1, 2, \ldots, p$ alcançado os $p$ diferentes autovetores de $\mathbf{S}_X$ relativo aos seus autovalores $\lambda_1, \lambda_2, \ldots, \lambda_p$, correspondentes.
Finalmente, a variância de cada Componente Principal será dada por $\text{Var}(\mathbf{u}_k\mathbf{X}) = \lambda_k$ para $k = 1, 2, \ldots, p$. Para encontrar os autovetores restantes usaremos o fato de que para cada $i \neq j$, $\mathbf{w}_i = \mathbf{u}_i^\mathsf{T}\mathbf{X}$ e $\mathbf{w}_i = \mathbf{u}_j^\mathsf{T}\mathbf{X}$ são ortogonais, então

$$\text{Var}(\mathbf{w}_i, \mathbf{w}_j) = \mathbb{E}[\mathbf{u}_i^\mathsf{T}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{u}_j] = \mathbf{u}_i^\mathsf{T}\mathbf{S}_X\mathbf{u}_j = 0.$$

## Componentes Principais: abordagem geométrica

Usando indução, assumimos que $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_{i-1}$ são os autovetores unitários de $\mathbf{S}_X$ associados com os maiores $i-1$ autovalores, e seja $\mathbf{u}_i$ o vetor que define o $i$-ésimo componente principal, $\mathbf{w}_i$. Então, $\mathbf{S}_X\mathbf{u}_j = \lambda_j\mathbf{u}_j$ para $j = 1, 2, \ldots, i-1$ e $\mathbf{u}_i^\top \mathbf{S}_X\mathbf{u}_j = \lambda_j\mathbf{u}_i^\top\mathbf{u}_j = 0$ para todo $j = 1, 2, \ldots, i-1$. Desde que $\lambda_j > 0$, temos que $\mathbf{u}_j^\top\mathbf{u}_j = 0$ para todo $j = 1, 2, \ldots, i-1$. Para computar $\mathbf{u}_i$, temos o Langrangiano

$$\mathcal{L}(\mathbf{S}, \mathbf{u}_i, \mathbf{u}_j, \lambda_i, \phi_j) = \mathbf{u}_i^\top\mathbf{S}_X\mathbf{u}_i - \lambda_i(\mathbf{u}_i^\top\mathbf{u}_i - 1) - \sum_{j=1}^{i-1}\phi_j\mathbf{u}_i^\top\mathbf{u}_j.$$

## Componentes Principais: abordagem geométrica

As condições necessárias para que $(\mathbf{u}_i, \lambda_i, \phi_1, \ldots, \phi_{j-1})$ seja máximo serão

$$2\mathbf{S}_X\mathbf{u}_i + \sum_{j=1}^{i-1} \phi_j \mathbf{u}_j = 2\lambda_i \mathbf{u}_i$$

com $\mathbf{u}_i^\mathsf{T}\mathbf{u}_i = 1$ e $\mathbf{u}_i^\mathsf{T}\mathbf{u}_j = 0$, para $j = 1, 2, \ldots, i-1$, de onde segue que para todo $j = 1, 2, \ldots, i-1$ teremos $\mathbf{u}_j^T\mathbf{S}_X\mathbf{u}_i + \frac{\phi_j}{2} = \lambda_j\mathbf{u}_j^\mathsf{T}\mathbf{u}_i + \frac{\phi_j}{2} = \lambda_i\mathbf{u}_j^\mathsf{T}\mathbf{u}_i$, e então $\phi_j = 2(\lambda_j - \lambda_i)\mathbf{u}_j^\mathsf{T}\mathbf{u}_i = 0$. Desde que o valor extremo associado é $\mathbf{u}_j^\mathsf{T}\mathbf{S}_X\mathbf{u}_i = \lambda_i = \mathrm{Var}(w_i)$, $u_i$ é autovetor de $\mathbf{S}_X$ restrito ao complemento ortogonal do span (Espaço gerado) $u_1, u_2, \ldots, u_{i-1}$.

# Componentes Principais: abordagem geométrica

Desde que o valor extremo associado é $\mathbf{u}_j^T \mathbf{S}_X \mathbf{u}_i = \lambda_i = \mathrm{Var}(w_i)$, $u_i$ é autovetor de $\mathbf{S}_X$ restrito ao complemento ortogonal do span (Espaço gerado) $u_1, u_2, \ldots, u_{i-1}$. Desde que os autovetores de $\mathbf{S}_X$ são distintos, $\mathbf{u}_i$ é o autovetor associado de $\mathbf{S}_X$ com o $i$-ésimo maior autovalor. Ainda, quando os autovalores de $\mathbf{S}_X$ são

distintos, cada autovetor $\mathbf{u}_i$ será único, e portanto serão os componentes principais de $\mathbf{X}$.

# Roteiro

# Em resumo: notação Johnson Wichern (2002)

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

$$\underbrace{\mathbf{Y}}_{(p \times p)} = \underbrace{\mathbf{W}^t}_{(p \times p)} \underbrace{\mathbf{X}}_{(p \times p)}; \tag{5}$$

# 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

$$\underbrace{\mathbf{Y}}_{(m\times p)} = \underbrace{\mathbf{W}^t}_{(m\times p)} \underbrace{\mathbf{R}}_{(p\times p)} ; \tag{6}$$

$$\underbrace{\mathbf{F}}_{(n\times m)} = \underbrace{\mathbf{X}}_{(n\times p)} \underbrace{\mathbf{W}}_{(p\times m)} . \tag{7}$$

# Em resumo: Propriedade 1

$$\mathbf{W}_j = \mathbf{u}_j^T\mathbf{X} = u_{j1}\mathbf{X}_1 + u_{j1}\mathbf{X}_2 + \cdots + u_{jp}\mathbf{X}_p \ , \ \forall \ j = 1, 2, \ldots, p.$$

Tal que

$$E(\mathbf{W}_j) = \mathbf{u}_j^T.$$

$$\mathrm{Var}(\mathbf{W}_j) = \mathbf{u}_j^T\boldsymbol{\Sigma}\mathbf{u}_j = \lambda_j \ , \ \forall \ j = 1, \ 2, \ \cdots, \ p.$$

$$\mathrm{Cov}(\mathbf{W}_j, \ \mathbf{W}_k) = \mathbf{u}_j^T\boldsymbol{\Sigma}\mathbf{u}_k = 0 \ , \ \forall \ j, k = 1, \ 2, \ \ldots, \ p \ \text{ e } \ j \neq k.$$

# Em resumo: Propriedade 2

A proporção da variância total de **X** que é explicada pela *j*-ésima componente principal é

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i} = \frac{\lambda_j}{\text{Traço}(\mathbf{S})}, \quad \forall\ j = 1,\ 2,\ \ldots,\ p.$$

# Análise de Componentes Principais (ACP):

## uma breve introdução

Rafael Pentiado Poerschke

Econometria II [1]

[1]Centro de Ciências Sociais e Humanas (CCSH)
Universidade Federal de Santa Maria (UFSM)

23 de março de 2024