

PCA - Análise de componentes principais

Juliano Cesar da Silva 770316

Universidade Federal de São Carlos

Fevereiro de 2024

① Introdução

② Componentes

③ Resultados

Introdução

Introdução

O que é um PCA ?

- Técnica estatística multivariada, diferente da análise de correlação (bivariada);
- Transforma o conjunto de variáveis originais em um conjunto de variáveis da mesma dimensão (Componentes principais).
- Um componente principal é uma combinação linear de todas as variáveis originais;
- É associada a redução de dimensionalidade perdendo o mínimo de informação possível;
- Os indivíduos são agrupados segundo suas variâncias.

Introdução

Objetivo

Aplicar tal técnica ao conjunto de dados *AP-2010-SP-REDUZIDO-REV02*

Conjunto de dados

V0011	AP0001	AP0062
3543402005001	10203	724
3543402005002	16785	1751
⋮	⋮	⋮
3543402005019	6461	578

- Com 19 variáveis e 903 linhas ao total

Componentes

Análise gráfica

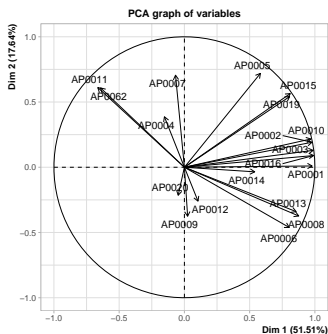


Figura 1: Conjunto de dados antigo

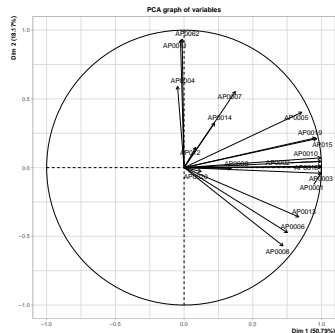


Figura 2: Conjunto de dados novo

Número de componentes ideais

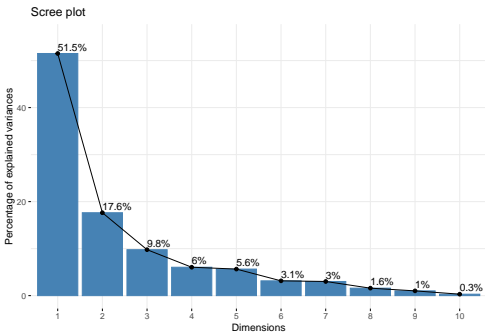


Figura 3: Conjunto de dados antigo

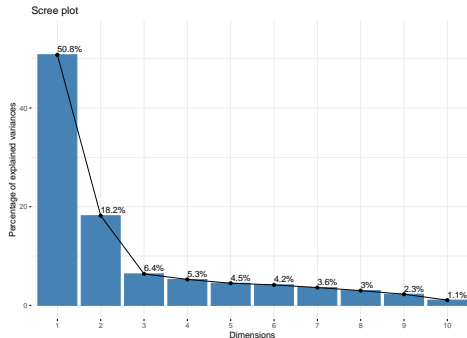


Figura 4: Conjunto de dados novo

Resultados

Resultados

Tabela dos autovalores e variância explicada

Componentes	Autovalores	(%) Variância explicada	(%) Variância cumulativa
1	9.65	50.79	50.79
2	3.45	18.17	68.96
3	1.21	6.38	75.35
4	0.99	5.26	80.61
5	0.85	4.49	85.11
:	:	:	:
19	0.00006	0.00003	100.00

Tabela de contribuição que cada variável adiciona ao componente

	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
AP0001	10.273994134				
AP0002	10.270123178				
AP0003	10.321621392				
AP0004		10.0957469281			
AP0005	7.602903315				
AP0006		6.4512670024			
AP0007		8.8571759233			
AP0008		9.3812475926			
AP0009					44.487399825
AP0010	10.240336981				
AP0011		24.9245067382			
AP012				63.9452782823	
AP0013	7.205821055				
AP0014					33.068808403
AP015	9.646218915				
AP0016	10.320660344				
AP0019	9.472490083				
AP0020			51.086915762		
AP0062		25.2993020051			

Figura 5: Tabela final

Componentes

Componente 1 (Total de residentes)

- AP0001 - Total de residentes do sexo masculino;
- AP0002 - Total de residentes do sexo feminino ;
- AP0003 - Total de residentes do (AP0001 + AP0002);
- AP0005 - Total de residentes autodeclarados brancos;
- AP0010 - Total de residentes com +- 5 anos que sabem ler/escrever;
- AP0013 - Total domicílios part. perm. tipo casa;
- AP0015 - Total de domicílios part. perm. por área de ponderação;
- AP0016 - Total de residentes por área de ponderação;
- AP0019 - Domicílios part. perm. ligados a rede geral de dist. [COM saneamento básico];

Componentes

Componente 2 (Etnia)

- AP0004 - Idade média calculada em anos e meses(V6033)
- AP0006 - Total de residentes autodeclarados pretos;
- AP0007 - Total de residentes autodeclarados amarelos;
- AP0008 - Total de residentes autodeclarados pardos;
- AP0011 - Total do rendimento nominal mensal dos domicílios part. perm [Renda da fam.];
- AP0062 - Valor média do rendimento nominal mensal dos dom. part. perm. com rendimento domiciliar [Renda fam. média].

Componentes

Componente 3 (Sem saneamento)

- AP0020 - Domicílios part. perm. abastecidos de outra forma [SEM saneamento básico];

Componentes

Componente 4 (Rendimento Barracos)

- AP0012 - Total do rendimento nominal mensal dos domicílios impro. [Renda da fam.(barracos)].

Componentes

Componente 5 (Domicílios Indígenas)

- AP0009 - Total de residentes autodeclarados indígenas;
- AP0014 - Total de dom. part. perm. do tipo casa em vila/condomínio