

01.Componentes Principais

Felipe Rocha

2022-11-30

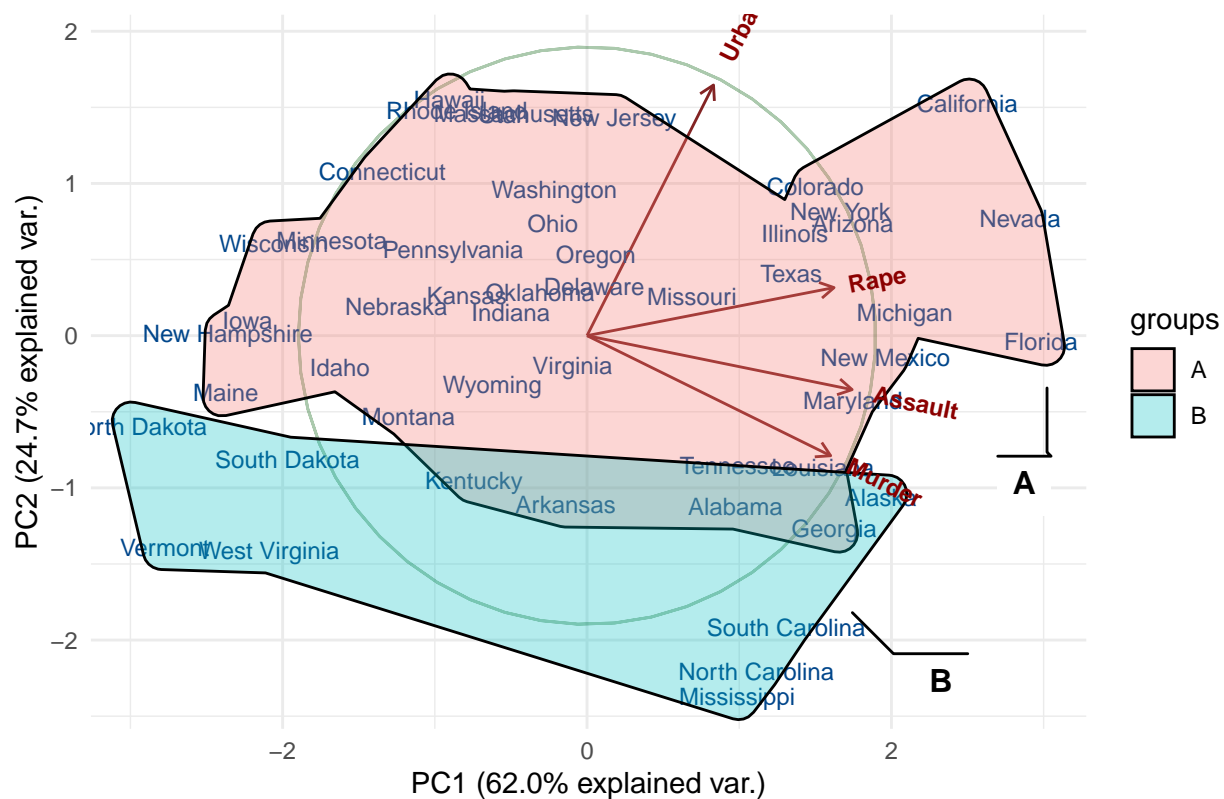
Antes de tudo, puxando a função de PCA

```
source("F:/Estudo/Estudo---UFV/UFV---GitHub/00.Functions/Function_PCA.R")
```

Exemplo 01: USArrests

10.2 Principal Components Analysis

```
pr.out=prcomp(USArrests, scale=TRUE)
pr.out$rotation=-pr.out$rotation
pr.out$x=-pr.out$x
ggbiplot(pcobj=pr.out, obs.scale = 1, var.scale = 1,
         groups = ifelse(USArrests$UrbanPop>=50,"A","B"),ellipse = TRUE,
         circle = TRUE,labels = rownames(USArrests) ) +
  scale_color_discrete(name = '') +
  theme(legend.direction = 'horizontal', legend.position = 'top')+theme_minimal()
```



Obs.: O groups é apenas para testar a opção 'groups' da função

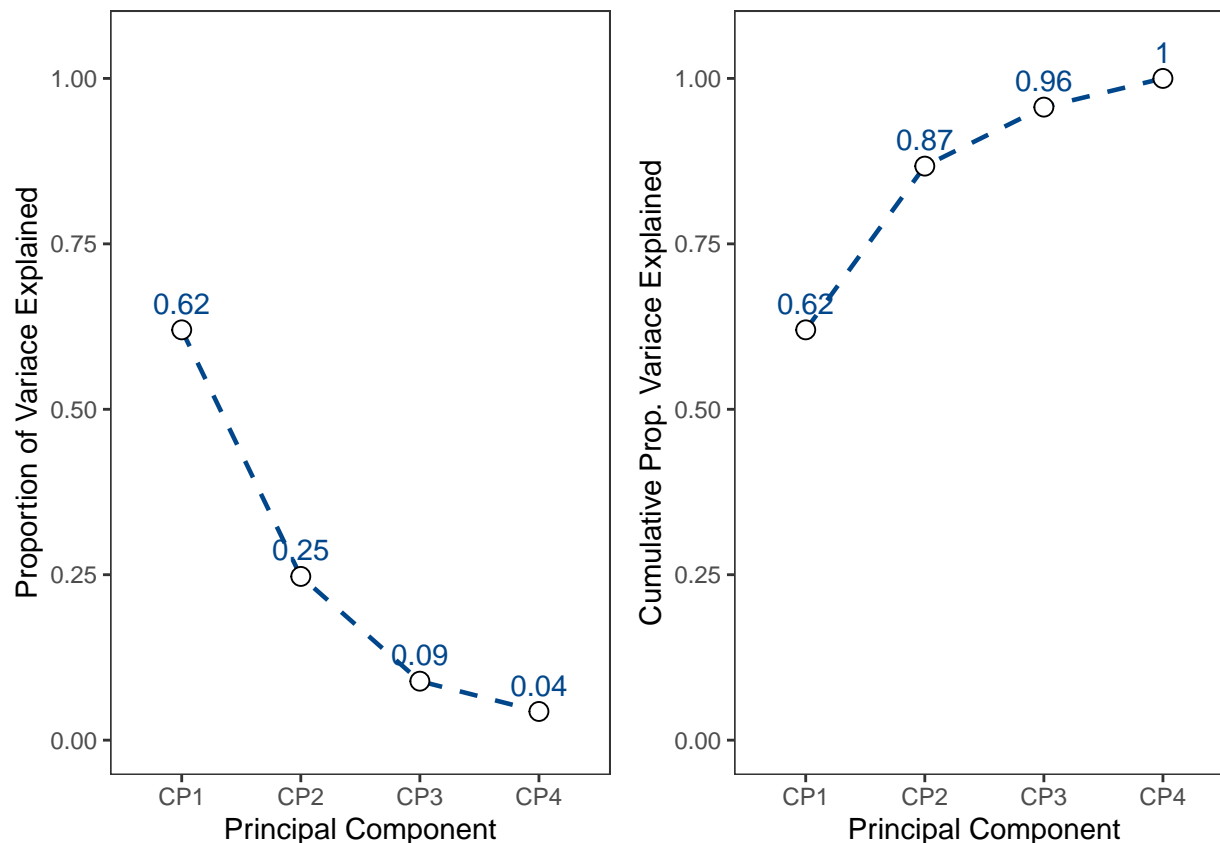
Na Figura, vemos que o **primeiro** vetor de carregamento coloca peso aproximadamente igual em Assalto, Assassinato e Estupro, com peso muito menor em UrbanPop. Portanto, esse componente corresponde aproximadamente a uma medida das taxas gerais de crimes graves. O **segundo** vetor de carregamento coloca a maior parte de seu peso no UrbanPop e muito menos peso nos outros três recursos. Assim, esse componente corresponde aproximadamente ao nível de urbanização do estado. No geral, vemos que as variáveis relacionadas ao crime (Assassinato, Assalto e Estupro) estão localizadas próximas umas das outras, e que a variável UrbanPop está distante das outras três. Isso indica que as variáveis relacionadas ao crime estão correlacionadas entre si – estados com altas taxas de homicídios tendem a ter altas taxas de agressão e estupro – e que a variável UrbanPop está menos correlacionada com as outras três.

Podemos examinar as diferenças entre os estados por meio dos dois vetores de pontuação dos componentes principais mostrados na Figura. Nossa discussão sobre os vetores de carga sugere que estados com grandes pontuações positivas no **primeiro componente**, como Califórnia, Nevada e Flórida, têm altas taxas de criminalidade, enquanto estados como Dakota do Norte, com pontuações negativas no **primeiro componente**, têm baixas taxas de criminalidade.

A Califórnia também tem uma pontuação alta no **segundo componente**, indicando um alto nível de urbanização, enquanto o oposto é verdadeiro para estados como o Mississippi. Estados próximos de zero em ambos os componentes, como Indiana, têm níveis aproximadamente médios de criminalidade e urbanização.

The Proportion of Variance Explained

```
PCA_Prop_Var(pr.out)
```



Nos dados do *USArrests*, o **primeiro componente principal** explica 62,0% da variação nos dados, e o próximo componente principal explica 24,7% da variação. Juntos, os dois primeiros componentes principais explicam quase 87% da variância dos dados, e os dois últimos componentes principais explicam apenas 13% da variância. Isso significa que a Figura fornece um resumo bastante preciso dos dados usando apenas duas dimensões. O PVE de cada componente principal, assim como o PVE cumulativo, é mostrado na Figura.

Nos dados do *USArrests*, o **primeiro componente principal** explica 62,0% da variação nos dados, e o próximo componente principal explica 24,7% da variação. Juntos, os dois primeiros componentes principais explicam quase 87% da variância dos dados, e os dois últimos componentes principais explicam apenas 13% da variância. Isso significa que a Figura fornece um resumo bastante preciso dos dados usando apenas duas dimensões. O PVE de cada componente principal, assim como o PVE cumulativo, é mostrado na Figura.

Avaliação referente ao conteúdo ministrado na Semana 1

Questão 1

Correto
Atingiu 1,00 de 1,00
🚩 Marcar questão

Marque as alternativas que considerar verdadeiras.

- ☐ a. O efeito da escala da variável não altera a solução de componentes principais.
- ☒ b. Os Componentes principais são variáveis latentes independentes.✖
- ☒ c. A técnica consiste em transformar um conjunto original de variáveis em outro de variáveis latentes.✔
- ☒ d. A solução de componentes principais é dada pela obtenção dos autovetores e autovalores da matriz de covariâncias (ou correlação) dos dados.✔
- ☒ e. Os primeiros componentes captam o maior percentual de variação dos dados.✔

Sua resposta está correta.

As respostas corretas são:

A técnica consiste em transformar um conjunto original de variáveis em outro de variáveis latentes,

A solução de componentes principais é dada pela obtenção dos autovetores e autovalores da matriz de covariâncias (ou correlação) dos dados,

Os primeiros componentes captam o maior percentual de variação dos dados.

Questão 2

Correto
Atingiu 1,00 de 1,00
🚩 Marcar questão

Não é um objeto do R.

- ☐ A. Funções
- ☐ B. listas
- ☒ C. Apply✔
- ☐ D. Matrizes
- ☐ E. Vetores

Sua resposta está correta.

A resposta correta é: Apply

Questão 3

Correto
Atingiu 1,00 de 1,00
🚩 Marcar questão

Verifique a afirmação a seguir:

"A solução dos componentes principais é única."

Escolha uma opção:

- ☐ Verdadeiro
- ☒ Falso ✔

A resposta correta é 'Falso'.

Questão 4

Correto
Atingiu 1,00 de 1,00
🚩 Marcar questão

Todos os pacotes necessários para realização das análises já estão disponíveis no software R após a instalação do mesmo.

Escolha uma opção:

- ☐ Verdadeiro
- ☒ Falso ✔

A resposta correta é 'Falso'.

**Questão
5**

Correto

Atingiu 1,00 de
1,00

🚩 Marcar
questão

Para realizar um aprendizado não supervisionado é necessário ter tanto um conjunto de variáveis independentes quanto uma variável dependente.

Escolha uma opção:

- ☐ Verdadeiro
- ☒ Falso ✓

A resposta correta é 'Falso'.