

# Avaliação 1 - Análise Multivariada 2

Tharso Monteiro

Análise de Componentes Principais  
Análise Fatorial

4 de novembro de 2021

O objetivo da ACP é a criação de um conjunto de variáveis chamadas componentes principais, que substituam o conjunto original enquanto retêm maior parte da informação contida nele.

Os componentes principais são não correlacionados entre si e são ordenados por critério de variância decrescente, isto é, os primeiros componentes apresentam mais informação que os subsequentes.

O primeiro componente principal,  $Z_1$ , de um conjunto de variáveis  $X_1, X_2, \dots, X_p$ , centradas e escaladas, é dado por

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p, \quad (1)$$

sujeito a:

$$\sum_{j=1}^p \phi_{j1}^2 = 1, \quad (2)$$

de maneira que  $Z_1$  tenha a maior variância possível.

De maneira similar,

O segundo componente principal,  $Z_2$ , de um conjunto de variáveis  $X_1, X_2, \dots, X_p$ , centradas e escaladas, é dado por

$$Z_2 = \phi_{12}X_1 + \phi_{22}X_2 + \dots + \phi_{p2}X_p, \quad (3)$$

sujeito à não-correlação com  $Z_1$  e a

$$\sum_{j=1}^p \phi_{j1}^2 = 1. \quad (4)$$

Os componentes subsequentes são calculados de maneira similar.



Figura: Imagem de satélite de Pyongyang, capital da RPDC.

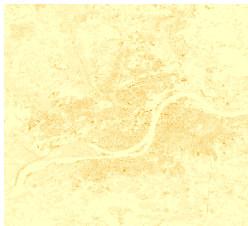


Figura: Decomposição  
em vermelho

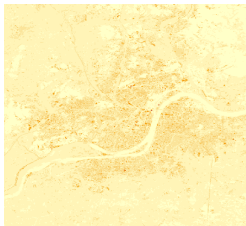


Figura: Decomposição  
em verde

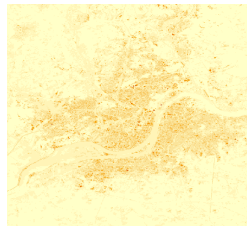


Figura: Decomposição  
em azul

Utilizando a função **prcomp** e reconstruindo a imagem a partir dos *principal component scores* e da matriz de rotação:

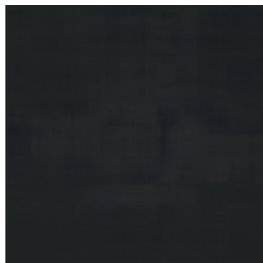


Figura: Recomposição da imagem com 5 CPs



Figura: Recomposição da imagem com 20 CPs



Figura: Recomposição da imagem com 40 CPs

A análise fatorial busca explicar a variância em um conjunto de variáveis mensuráveis com base em variáveis latentes, isto é, conceitos que não podem ser medidos de maneira direta, mas que por suposição, são relacionados às variáveis mensuráveis.

O modelo da análise fatorial é baseado na regressão múltipla. Entretanto, diferentemente da regressão múltipla, as variáveis de interesse são modeladas com base em variáveis latentes. Por isso, a estimação direta dos coeficientes associados às mesmas não é possível. Entretanto, existem métodos de estimação indireta e de ajuste dos valores desses parâmetros de maneira que a interpretação dos fatores correspondentes seja mais intuitiva.



Sejam  $x_{red}$ ,  $x_{green}$  e  $x_{blue}$  as variáveis que correspondem ao valor de um determinado pixel na decomposição RGB.

Assumimos o modelo com uma variável latente  $f$ .

$$x_{red} = \lambda_1 f + u_1$$

$$x_{green} = \lambda_2 f + u_2$$

$$x_{blue} = \lambda_3 f + u_3$$

Utilizando a função **factanal** da linguagem R, que utiliza o método de máxima verossimilhança e rotação Varimax, obtivemos os seguintes resultados:

$$\hat{\lambda}_1 = 0.978$$

$$\hat{\lambda}_2 = 0.986$$

$$\hat{\lambda}_3 = 0.936$$

$$\hat{\sigma}_{u_1}^2 = 0.044$$

$$\hat{\sigma}_{u_2}^2 = 0.028$$

$$\hat{\sigma}_{u_3}^2 = 0.124$$

Podemos interpretar  $f$  como sendo a captação de luz branca pelo satélite.