

Trabalho Multivariada

Thiago Tavares Lopes

2025-05-07

Análise de Componentes Principais (PCA)

A análise de componentes principais foi realizada sobre quatro variáveis quantitativas (pH , N , $Oxig$ e M_Org), utilizando a matriz de correlação. A Tabela 1 apresenta os autovalores e a variância explicada por cada componente.

Table 1: Autovalores e variância explicada

Componente	Autovalor	% Variância Explicada	% Acumulada
Dimensão 1	2.63	65.78%	65.78%
Dimensão 2	0.90	22.58%	88.36%
Dimensão 3	0.33	8.24%	96.60%
Dimensão 4	0.14	3.40%	100.00%

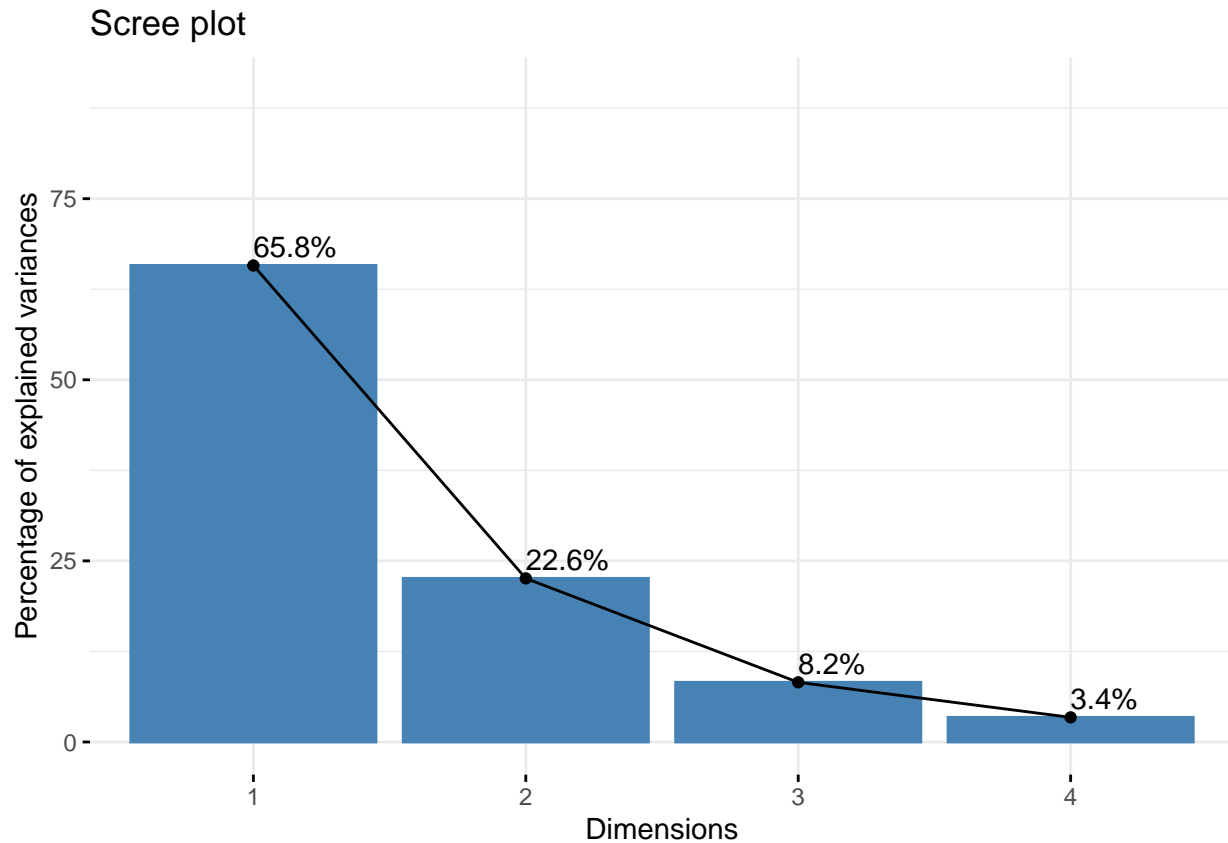
Observa-se que as duas primeiras componentes principais explicam juntas aproximadamente 88,36% da variância total dos dados, o que indica que uma boa parte da informação original pode ser representada em um espaço bidimensional.

A Tabela 2 mostra a correlação das variáveis originais com os componentes principais. A Dimensão 1 apresenta forte correlação positiva com $M.Org$ ($r = 0,94$), pH ($r = 0,90$) e N ($r = 0,86$), enquanto a Dimensão 2 se associa fortemente à variável $Oxig$ ($r = 0,88$).

Table 2: Correlação das variáveis com os componentes principais

Variável	Dim.1	Dim.2	Dim.3	Dim.4
pH	0.90	-0.03	0.42	0.15
N	0.86	0.33	-0.36	0.16
Oxig	-0.45	0.88	0.15	-0.01
M_Org	0.94	0.15	0.00	-0.29

Com base nesses resultados, conclui-se que a primeira dimensão representa um gradiente de qualidade ambiental relacionado a acidez e matéria orgânica, enquanto a segunda está mais associada à presença de oxigênio dissolvido.

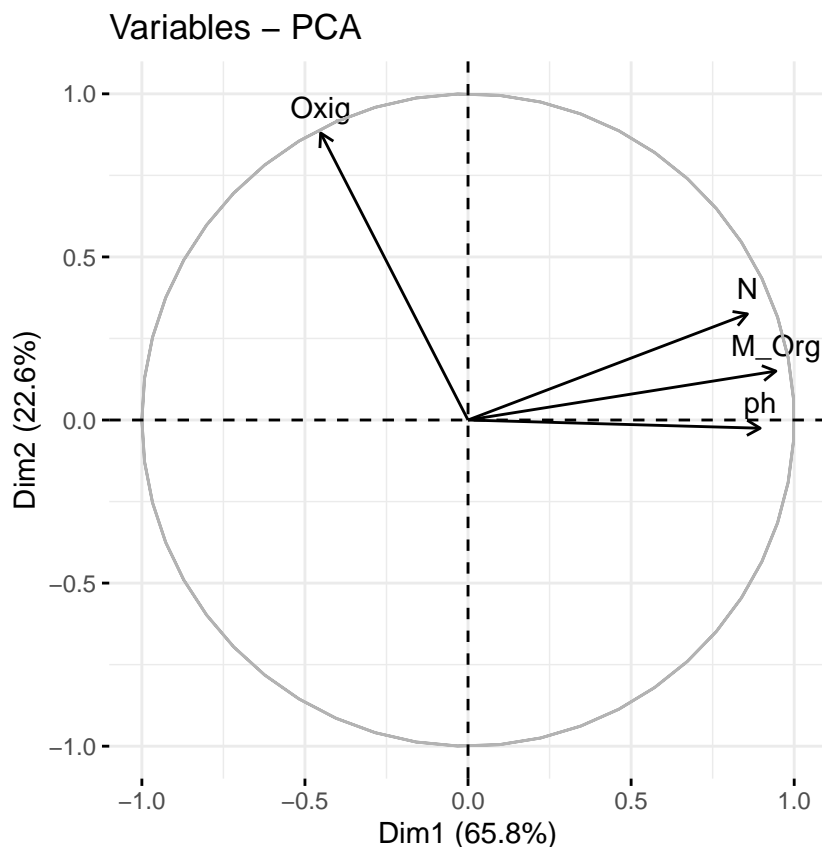


Análise do Gráfico de Scree Plot

O gráfico de scree plot apresentado na Figura~?? exibe a proporção da variância explicada por cada componente principal (PC) obtido na análise.

- A **Dimensão 1** explica **65,8%** da variância total dos dados, indicando que uma grande parte da estrutura dos dados é capturada por esse componente.
- A **Dimensão 2** adiciona mais **22,6%** à explicação, totalizando aproximadamente **88,4%** da variância explicada pelas duas primeiras componentes.
- As **Dimensões 3 e 4** contribuem com apenas **8,2%** e **3,4%**, respectivamente, o que evidencia uma queda acentuada após a segunda componente.

Essa queda abrupta na variância explicada no gráfico em questão, sugere que os dois primeiros componentes são suficientes para representar a maior parte da informação dos dados.



Análise do Gráfico de Correlação das Variáveis

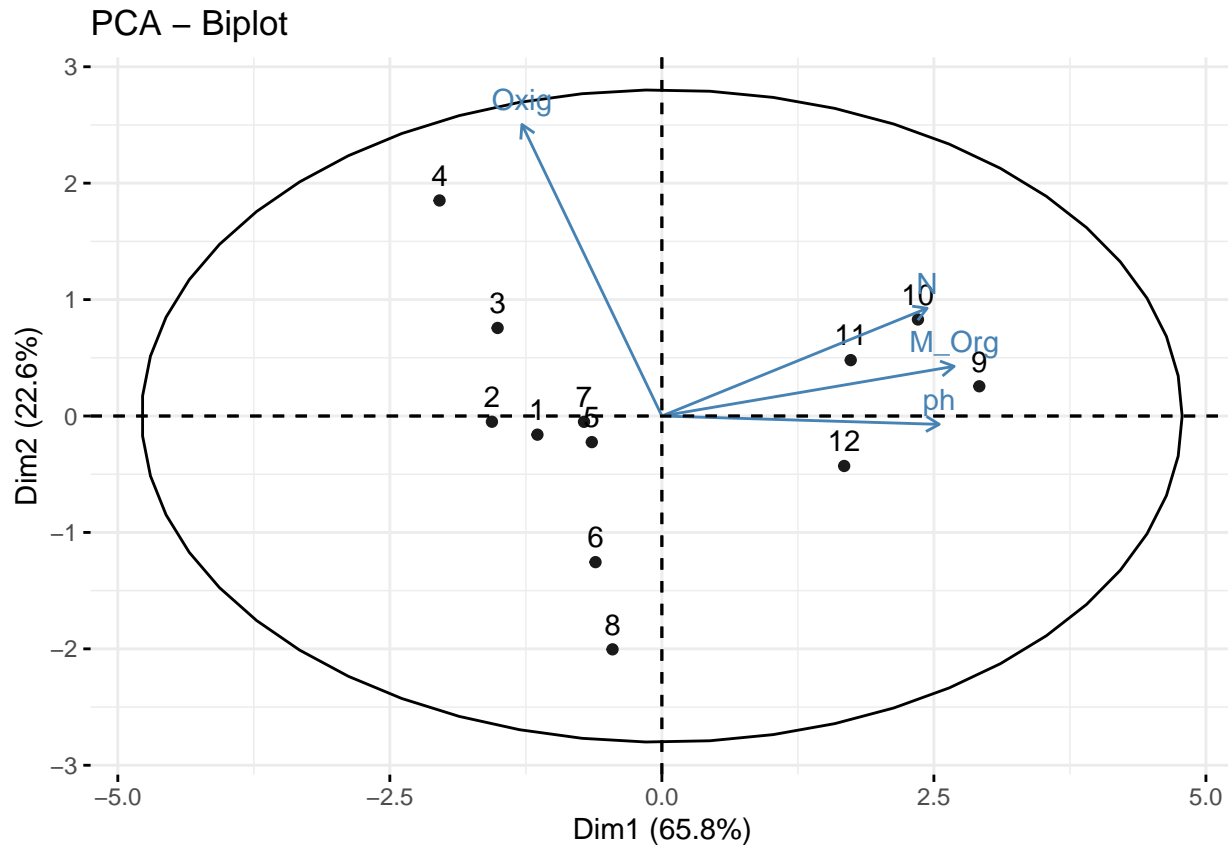
A Figura apresenta o gráfico de correlação que mostra a projeção das variáveis originais sobre o plano definido pelas duas primeiras componentes principais (Dim.1 e Dim.2), responsáveis por 88,4% da variância total dos dados (65,8% e 22,6%, respectivamente).

Interpretação dos eixos:

- O eixo horizontal (**Dim.1**) está fortemente associado às variáveis *pH*, *N* e *M_Org*, representando um gradiente relacionado à acidez da água e à presença de matéria orgânica.
- O eixo vertical (**Dim.2**) está altamente correlacionado com a variável *Oxig*, indicando uma direção de variação independente das demais variáveis.

Distribuição e relações entre as variáveis:

- As variáveis *pH*, *N* e *M_Org* aparecem agrupadas e apontam em direção semelhante, com ângulos pequenos entre si, sugerindo correlação positiva mútua e associação com a Dim.1.
- A variável *Oxig* está orientada aproximadamente na direção da Dim.2, com ângulo obtuso em relação às demais variáveis, o que indica correlação negativa moderada com elas e forte contribuição para a segunda dimensão.



Análise do Biplot da Análise de Componentes Principais (PCA)

A Figura apresenta o *biplot* da Análise de Componentes Principais (PCA), no qual são representadas simultaneamente as observações (pontos numerados de 1 a 12) e as variáveis originais (*pH*, *N*, *Oxi*, *M_Org*) projetadas no plano formado pelas duas primeiras componentes principais.

As duas primeiras componentes principais explicam juntas **88,4%** da variância total dos dados (**65,8%** para a Dim.1 e **22,6%** para a Dim.2), permitindo uma representação adequada dos dados em duas dimensões.

Interpretação das variáveis:

- As variáveis *pH*, *N* e *M_Org* estão fortemente associadas à Dimensão 1, apontando para a direita, o que indica correlação positiva com esse eixo.
- A variável *Oxi* está direcionada para cima, caracterizando forte correlação com a Dimensão 2 e pouca relação com a Dimensão 1, o que sugere independência em relação às demais variáveis.

Interpretação das observações:

- As observações **9, 10 e 11** encontram-se na extremidade direita do gráfico, indicando maiores valores de *pH*, *N* e *M_Org*, possivelmente associadas ao ambiente **urbano**.
- As observações **1 a 4** estão mais à esquerda, apresentando baixos valores dessas variáveis, características típicas da **nascente**.
- A observação **8**, posicionada na parte inferior do gráfico, apresenta baixa contribuição de todas as variáveis, podendo indicar um perfil atípico, possivelmente relacionado ao ambiente **rural**.

Conclusão: o biplot permite observar claramente a separação das amostras com base nas características físico-químicas da água. As amostras urbanas tendem a se agrupar em uma região de altos valores de *pH*,

nitrogênio e matéria orgânica, enquanto as amostras de nascente se distinguem por concentrações menores. A variável *Oxig* se destaca como fator adicional de diferenciação no eixo vertical.

