

# Análise de componentes principais

**Marília Melo Favalesso**

**Thaís Maylin Sobjak**

Mestrandas do Programa de Conservação e Manejo de Ambientes Naturais



# Ordenação

- Simplificação de dados multivariados/redução de dados multivariados
- Começa com  $n$  variáveis gerando um número menor de variáveis
- Também pode ser usada para separar/discriminar amostras

**Os ecólogos e cientistas de maneira geral costumam utilizar cinco tipos de ferramentas:**

Análise de componentes principais (PCA)\*

Análise fatorial

Análise de correspondência

Análise de coordenadas principais

Escalonamento multidimensional não métrico

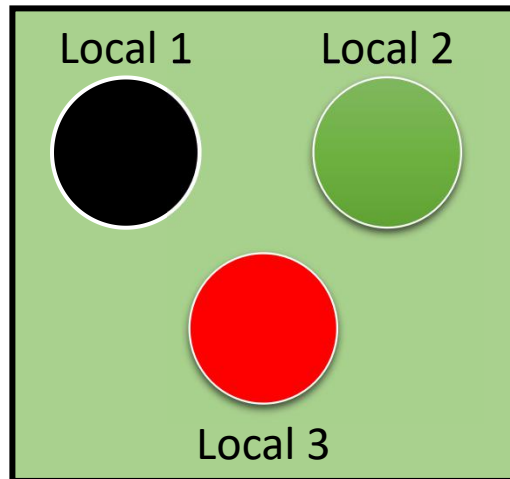


# Análise de componentes principais (PCA)

## **Objetivos:**

- Redução das variáveis originais em um número menor de variáveis com o intuito de facilitar a visualização dos dados e assim evidenciar a relação entre as observações realizadas e as variáveis amostradas
- As novas variáveis não são correlacionadas
- Facilitar a interpretação pela descoberta de relacionamentos não suspeitos previamente
- As novas variáveis podem ser usada na ANOVA e em regressões logísticas
- Criar índices ecológicos

# Conceito



3 Locais amostrados  
12 parcelas/local  
Abundância de 3 espécies



Resultados da amostragem

Locais	Parcelas	Espécie 1	Espécie 2	Espécie 3
1	1	2	1	3
1	2	3	2	4
1	3	4	3	3
1	4	3	2	2
1	5	2	1	1
1	6	1	2	2
1	7	2	3	2
1	8	3	4	1
1	9	2	3	2
1	10	1	2	1
1	11	4	3	1
1	12	5	5	2
2	1	6	2	3
2	2	3	3	4
2	3	4	4	2
2	4	5	5	1
...	...	...	...	...

# Conceito

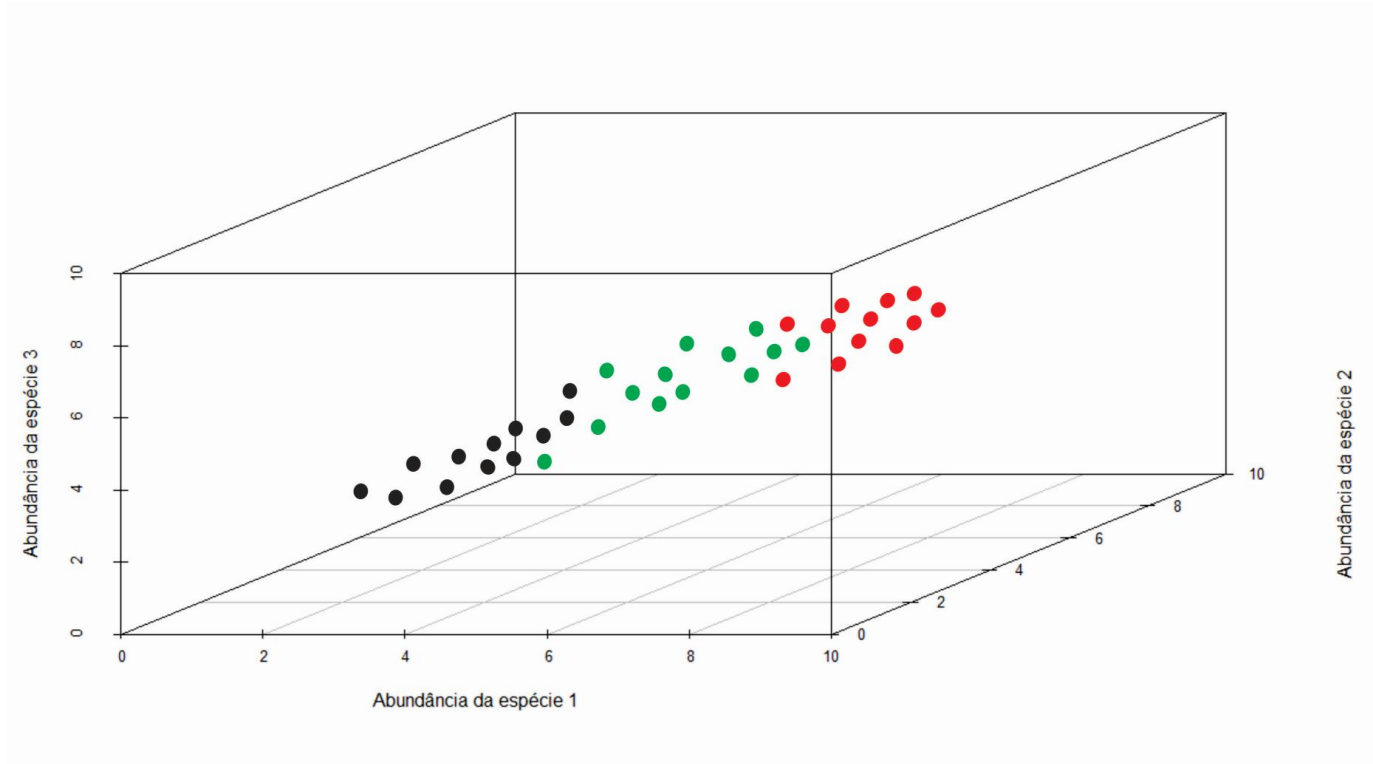
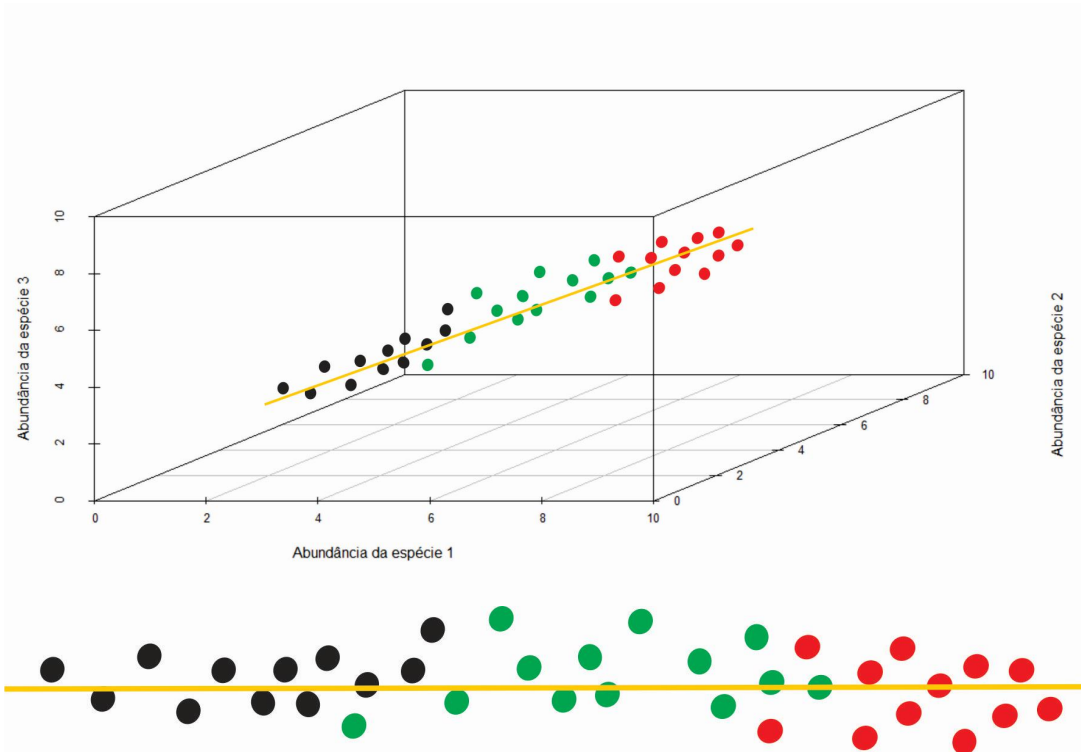


Gráfico tridimensional das variáveis amostradas

Matriz de correlação de Pearson para as variáveis amostradas

	Espécie 1	Espécie 2	Espécie 3
Espécie 1	1	0,81	0,95
Espécie 2	0,81	1	0,79
Espécie 3	0,95	0,79	1

# Conceito



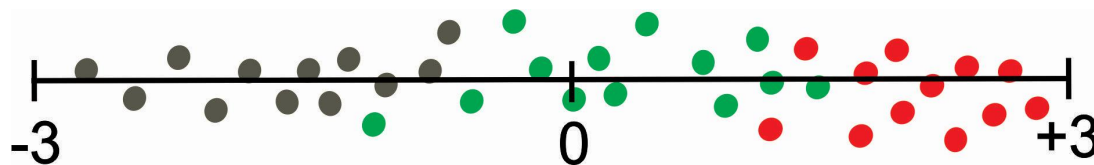
**Componente principal 1 (abundância de espécies)**

**Essa é a ideia da PCA:**

Encontrar um número menor de dimensões que explique uma quantidade suficiente de informações das variáveis originais (contendo grande parte da variação original dos dados).

Retem a maior parte da variação dos dados (autovalores)

# Conceito



**Escores dos componentes principais**

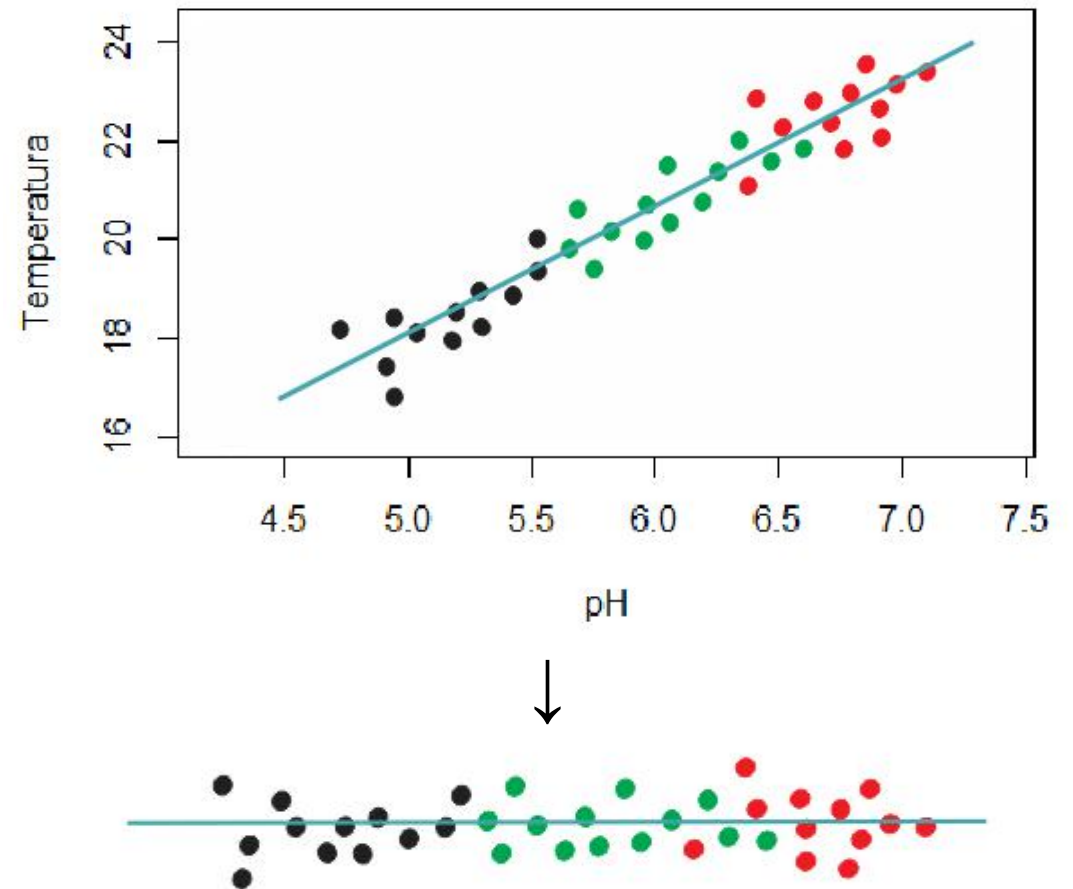
Os dados multivariados originais  
apresentavam 3 dimensões, agora eles  
estão resumidos como uma só (um único  
valor ao invés de três)

# Conceito

E se também fossem coletados dados ábóticos locais?

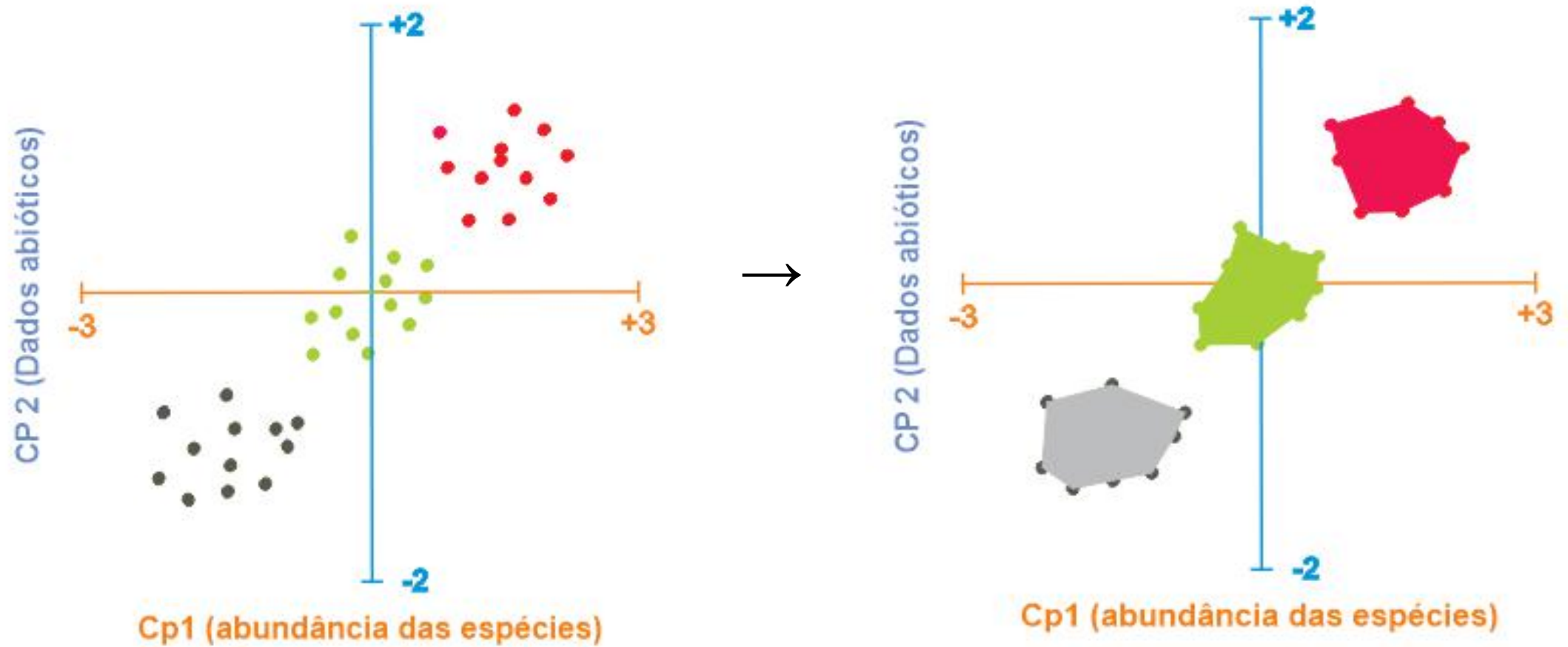
Locais	Parcela	Temperatura	pH
1	a	6.37	6.37
1	b	7.11	7.11
1	c	5.75	5.75
1	d	5.09	5.09
1	e	4.63	4.63
1	f	4.2	4.2
1	g	5	5
1	h	7.41	7.41
1	i	7	7
1	j	5.55	5.55
2	a	7.07	7.07
2	b	7.32	7.32
2	c	6.39	6.39
2	d	5.79	5.79
2	e	5.94	5.94
2	f	5.25	6.37
...	...	...	...

Correlação = 0,90





# Conceito



Resultado visual da PCA

# Pressupostos

- Apenas variáveis quantitativas - não se deve incluir qualitativa (análise de correspondência)
- Idealmente o número de observações ( $n$ ) > 50 ou 5 vezes mais observações do que número de variáveis (existem outros autores que discordam)
- Correlação multivariada entre as variáveis (teste de esfericidade de Bartlett)
- Índice KMO > 0,70
- Normalidade multivariada
- Padronização das variáveis em mesma escala - evitar que os eixos da PCA não sejam dominado por variáveis com grandes unidades de medida
- Autovalor > 1

# Alguns exemplos

No R

# Interpretação final

