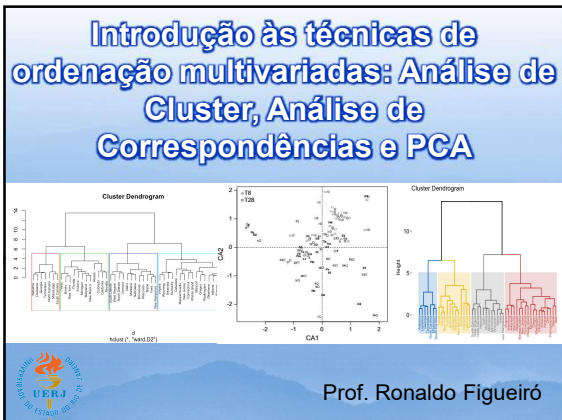
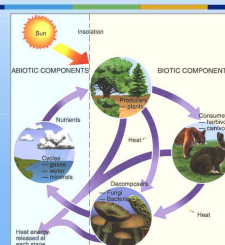




O ambiente e suas interações

■ O ambiente de qualquer organismo inclui os seguintes componentes:

- **Abiótico:** características do meio físico
- **Biótico:** a comunidade biológica



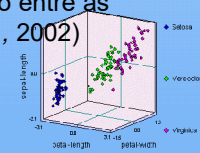
A complexidade das matrizes de dados de comunidades

- Inventários bióticos tradicionalmente resultam em matrizes onde se registram a presença ou abundância de cada espécie em um dado local ou ambiente
 - Dezenas de espécies e sítios
 - Variáveis ambientais



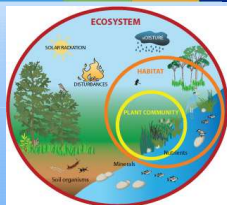
O que são análises multivariadas?

- As análises multivariadas objetivam reduzir um grande número de variáveis a poucas dimensões com o mínimo de perda de informação, permitindo a detecção dos principais padrões de similaridade, de associação e de correlação entre as variáveis (Lewinsohn *et al.*, 2002)



Diferenças Básicas para os testes de hipóteses univariados:

- NÃO há necessidade de hipóteses estabelecidas *a priori*.
- Natureza exploratória
- Várias variáveis INDEPENDENTES.



Entendendo dados multidimensionais

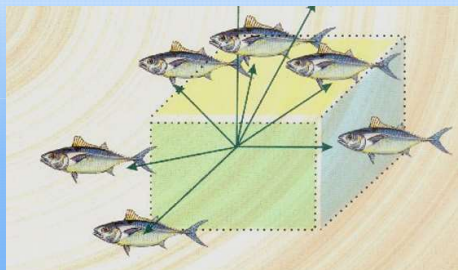
- Objetos e Descritores
 - Amostras / Espécies / dados ambientais

	B1	B2	B3
X1	X ₁₁	X ₁₂	X ₁₃
X2	X ₂₁	X ₂₂	X ₂₃
X3	X ₃₁	X ₃₂	X ₃₃
X4	X ₄₁	X ₄₂	X ₄₃
X5	X ₅₁	X ₅₂	X ₅₃



Exemplo retirado de Valentin (2012)

Representação vetorial



Tipos de modo de análise

- Matriz de semelhança entre amostras -> Modo Q

	sp 1	sp 2	profundidade	Temperatura
Estação A	X ₁₁	X ₁₂	X ₁₃	X ₁₄
Estação B	X ₂₁	X ₂₂	X ₂₃	X ₂₄
Estação C	X ₃₁	X ₃₂	X ₃₃	X ₃₄

- Matriz de semelhança entre descritores -> Modo R

	sp 1	sp 2	profundidade	Temperatura
Estação A	X ₁₁	X ₁₂	X ₁₃	X ₁₄
Estação B	X ₂₁	X ₂₂	X ₂₃	X ₂₄
Estação C	X ₃₁	X ₃₂	X ₃₃	X ₃₄

Tipos de dados

- Quantitativos (a)
- Semi-quantitativos (b)
- Qualitativos (c)

(a)		(b)		(c)		
sp A	sp B	Células / mL	Código	Estação	Calcário	Cascalho
34	5	0 - 10	1	Estação A	0	1
12	42	10 - 100	2	Estação B	0	0
5	80	100 - 1000	3	Estação C	1	0
65	4	1000 - 10000	4	Estação D	0	1

Medidas de semelhança

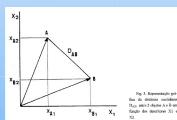
- Associação de objetos (Modo Q)

Coefficiente de similaridade	Fórmula	Varição
Jaccard	$a / a + b + c$	0 a 1
Sorensen	$2a / 2a + b + c$	0 a 1

A = número de espécies em comum aos dois objetos
 B = número de espécies exclusivas de um objeto
 C = número de espécies exclusivas do outro objeto

Medidas de distância

- Distância Euclidiana



$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

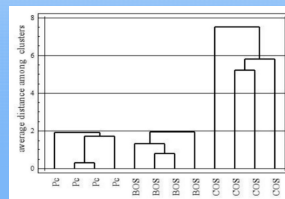
- Distância de Bray-Curtis

- Varia entre 0 (similaridade) e 1 (dissimilaridade)
- Fortemente influenciado por espécies dominantes

$$BC_{ij} = \frac{2C_{ij}}{S_i + S_j}$$

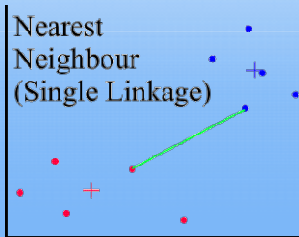
Análise de Cluster

- Útil para se estabelecer similaridade entre sítios de coleta, comunidades, etc.
- Natureza principalmente gráfica: gera um dendrograma, similar a um cladograma.



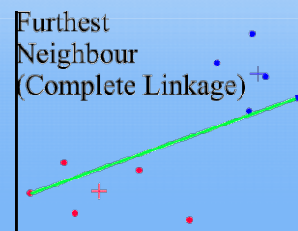
Método por ligação simples

- Vizinho mais próximo -> o dendrograma é montado a partir dos objetos mais similares



Método por ligação completa

- Vizinho mais distante
- Um elemento só se funde a um grupo se for ligado a todos os elementos desse grupo



Método de Ward

- Um grupo será fusionado a outro se este proporcionar o menor aumento da variância intragrupo

Hierarchical Grouping to Optimize an Objective Function

Joe H. Ward, Jr.

Journal of the American Statistical Association, Vol. 58, No. 301 (Mar., 1963), 236-244.

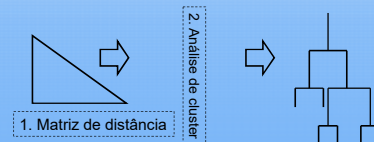
Stable URL:
[http://links.jstor.org/sici?sici=0162-1459\(196303\)29:58%3A301%3C236%3AHGTOAO%3E2.0.CO%3B2-9](http://links.jstor.org/sici?sici=0162-1459(196303)29:58%3A301%3C236%3AHGTOAO%3E2.0.CO%3B2-9)

Journal of the American Statistical Association is currently published by American Statistical Association.

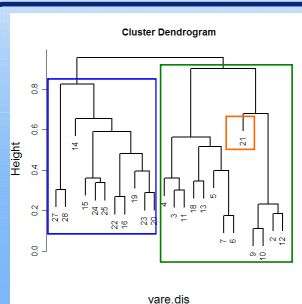


Análise de Cluster

A análise de Cluster tem como objetivo particionar o conjunto original de dados em subconjuntos menores (clusters) que guardam algumas características semelhantes entre si com base em medidas de distância.



Classificação hierárquica

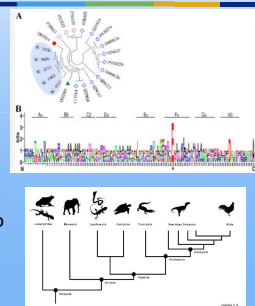


Aumento da
dissimilaridade

Classificação hierárquica

- Principais aplicações:

- Filogenia (métodos antigos)
- Bioinformática
 - Busca de grupos homólogos em famílias de genes
 - Algoritmos de clusterização utilizados em plataformas de genotipagem para atribuir automaticamente genótipos



Exemplo em artigos científicos:

Mem Inst Oswaldo Cruz, Rio de Janeiro, Vol. 102(8): 895-900, December 2007 895

Association of insecticide use and alteration on *Aedes aegypti* susceptibility status

Maria de Lourdes da Graça Macoris^{1,2}, Maria Teresa Macoris Andrighetti,
Vanessa Camargo Garbeloto Otera, Lídia Raquel de Carvalho*,
Antonio Luiz Caldas Júnior*, William G Brogdon***

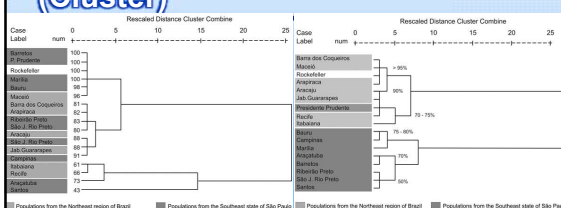
Superintendência de Controle de Endemias, SUCEN, Núcleo de Pesquisa SR-11/Marília, Av. Santo Antonio 1627, 17506-040 Marília, São Paulo, Brasil *Instituto de Biotecnologia **Faculdade de Medicina, Universidade Estadual Paulista, Botucatu, São Paulo, Brasil ***Centers for Disease Control and Prevention, Atlanta GA, USA

Objetivos:

- Comparar os níveis de resistência encontrados em *Aedes aegypti* de quatro estados brasileiros com diferentes históricos de uso de inseticidas:
 - São Paulo, Alagoas, Sergipe, Pernambuco



Análise de agrupamentos (Cluster)



- RR 95
- Rockfeller -> cepa suscetível
- Larvas (Esquerda)
 - Itabaiana, Recife, Araçatuba e Santos -> menos suscetíveis
- Adultos (direita)
 - Bauru, Campinas, Marília, Araçatuba, Barretos, Ribeirão Preto, S.J. Rio Preto, Santos -> menos suscetíveis

Exemplo em artigos científicos:

REVISTA DE BIOLOGIA TROPICAL

Water pollution and distribution of the black fly (Diptera: Simuliidae) in the Atlantic Forest, Brazil

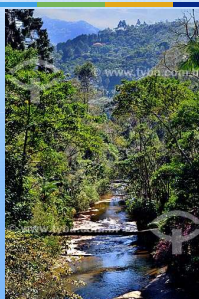
Tatiana N. Docile^{1,2,3}, Ronaldo Figueiro^{3,4}, Leonardo H. Gil-Azevedo⁵ & Jorge L. Nessimiani¹

1. Laboratório de Entomologia, Departamento de Zoologia, Universidade Federal do Rio de Janeiro (UFRJ), Av. Carlos Chagas Filho, 373, Edifício do Centro de Ciências da Saúde Bloco A, sala A1-107, Ilha do Fundão, Rio de Janeiro, RJ, Brasil - CEP: 21941-902, Caixa Postal: 68044; tatdocile@gmail.com; jnessimiani@gmail.com
2. Programa de Pós-Graduação em Ecologia da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil.
3. Laboratório de Biotecnologia Ambiental, Centro Universitário Estadual da Zona Oeste (UEZO), Rio de Janeiro, Brasil; ronaldofigueiro@gmail.com
4. Centro Universitário de Volta Redonda (UniFOA), Volta Redonda, Brasil.
5. Departamento de Entomologia, Museu Nacional, Universidade Federal do Rio de Janeiro (UFRJ) Rio de Janeiro, Brasil; lugaazevedo@gmail.com

Received 03-X-2014. Corrected 25-II-2015. Accepted 27-III-2015.

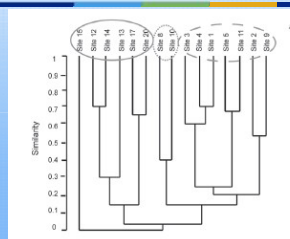
Objetivos:

- Avaliar a similaridade da composição de espécies entre rios previamente estabelecidos como preservados, intermediários e impactados



Uso da Análise de Cluster:

- Dados bióticos:
 - sítios 8 e 10 (rios impactados),
 - sítios 1 ao 5, 9 e 11, (rios intermediários)
 - e sítios 12 ao 15 e 17 (rios preservados)



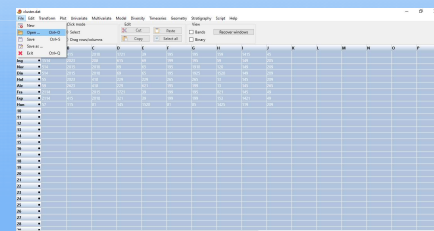
Estudo de caso no PAST:

- Foi realizado um levantamento dentre as populações de 7 países distintos sobre a preferência de cores de vestimenta. Quais populações tem perfis mais similares de preferência?

	1 - Branco	2 - Preto	3 - Verde	4 - Amarelo	5 - Azul	6 - Vermelho	7 - Rosa	8 - Marrom	9 - Cinza	10 - Laranja	11 - Roxo
1 - Fra	218	1621	126	225	113	121	1815	225	131	1618	
2 - Ing	238	1621	212	718	255	1518	169	185	218	212	
3 - Nor	822	69	212	718	721	1518	1225	1815	218	1922	
4 - Hol	231	161	212	718	75	199	1815	1815	218	2623	
5 - Ale	235	229	212	718	75	116	145	1815	218	193	
6 - Fra	212	1615	212	225	101	1518	1815	1815	218	1415	
7 - Esp	212	1621	126	225	113	141	1815	1315	1415		
8											
9											

Exemplo 1:

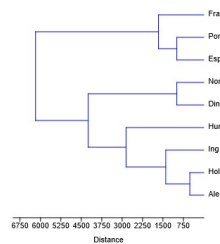
- No site ronaldofigueiro.wixsite.com/ecologia de vetores, na seção downloads, clique no item "Seminários de Verão" e baixe o arquivo cluster.dat e abra pelo PAST.



Exemplo 1:

- Em "edit" -> marcar "select all"
- Em "Multivariate" -> selecionar "Classical clustering"
- Marcar "Ward Method" (1)
- Marcar "Rotate dendrogram" (2)
- Por último apertar o botão "compute" (3)

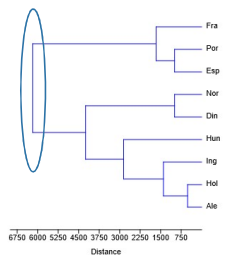
Resultados:



- O que o gráfico nos diz? Os valores indicados no eixo X são de distância euclidiana, ou seja, o quanto os grupos diferem entre si.

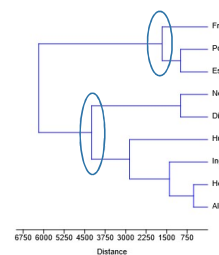
Resultados:

- Primeira conclusão: temos um grupo formado pelos latinos, e outro formado pelos anglo-saxões e nórdicos.



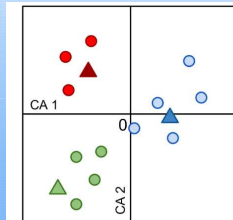
Resultados:

- Segunda conclusão: Os países mais similares entre si são Portugal e Espanha, no grupo latino, e Holanda e Alemanha, no grupo anglo-saxão/Nórdico.



Análise de correspondências

- Também conhecida como RA (reciprocate averaging), devido a se basear num princípio de médias recíprocas, que faz com que as colunas e as linhas tenham igual peso na análise.
- Cada dado é transformado em probabilidade
- Ordenação de espécies e amostras ocorre simultaneamente -> Modo R e Q
- Não exige normalidade



Exemplo em artigos científicos:

Local distribution of blackfly (Diptera, Simuliidae) larvae in two adjacent streams: the role of water current velocity in the diversity of blackfly larvae

Ronaldo Figueiró^{1,2*}, Érika Silva do Nascimento¹, Leonardo Henrique Gil-Azevedo¹, Marilza Maia-Herzog³ & Ricardo Ferreira Monteiro³

¹Laboratório de Simuliídeos e Oculocoronos, Referência Nacional em Simuliídeos, Oculocoronos e Macrotrichia, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Avenida Brasil, 4565, Mangueiras, 21045-900 Rio de Janeiro-RJ, Brasil.
²Laboratório de Ecologia de Insetos, Departamento de Ecologia, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Av. Brigadeiro Trompowski, s/n, CCS, Ilha do Fundão 21941-900 Rio de Janeiro-RJ, Brasil.
³Programa de Pós-Graduação em Ecologia (PPGE), Universidade Federal do Rio de Janeiro, Brasil.

ABSTRACT. Local distribution of blackfly (Diptera, Simuliidae) larvae on two adjacent streams: The role of water current velocity in the diversity of blackfly larvae. This study aimed to evaluate the influence of water velocity speed on the local distribution and composition of blackfly larvae. The larvae were collected from two adjacent streams.

Exemplo em artigos científicos:

- Objetivo: Identificar os padrões de preferência de microhabitat para larvas de simuliídeos
- Métodos: Quadrats distribuídos aleatoriamente por trechos de rios, tendo cada um sua composição de espécies e sua velocidade determinados

Exemplo em artigos científicos:

- *S. pertinax*
 - 0.95 – 1.2 e 1.2 – 1.45 (Rio A)
 - 0.76 – 0.92 e 0.92 – 1.0 (Rio B)
- *S. subpallidum*
 - 0.69 – 0.95 (Rio A)
 - 0.92 – 1.0 (Rio B)
- *S. (Inaequalium) sp.*
 - 0.4 – 0.6 e 0.6 e 0.7

- Faixas de velocidade com composições similares
 - 0.69 – 0.95, 0.95 – 1.2; 1.2 -1.5

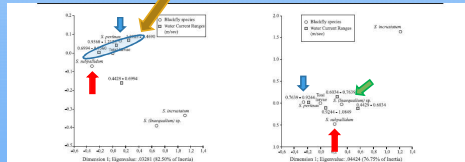
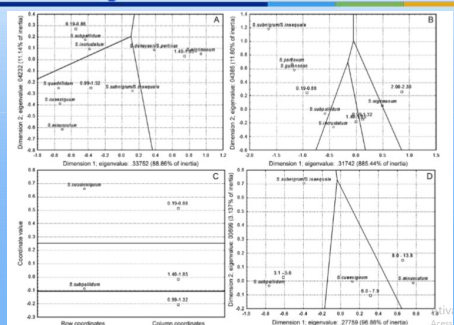


Fig. 2. Correspondence analysis for blackfly species and water current speed classes in stream A.

Fig. 3. Correspondence analysis for blackfly species and water current velocity classes in stream B.

Linha de Voronoi (não disponível no PAST)



Estudo de caso no Past

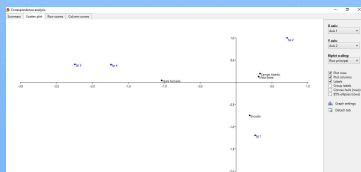
- Um pesquisador investigou a distribuição de quatro espécies de cactos por quatro fitofisionomias diferentes (arquivo correspondencia.dat)

	Sp 1	Sp 2	Sp 3	Sp 4
Mata baixa	67	97	5	11
Mata fechada	20	9	20	25
Encosta	35	7	1	2
Campo Aberto	12	20	1	2



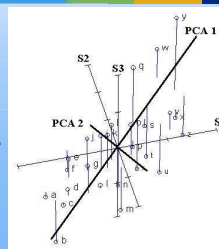
Resultados:

- Sp 2 associada aos ambientes de campo aberto e mata
- Sp 3 e Sp 4 associadas à Mata fechada
- Sp 1 associada ao ambiente de encosta
- Campo Aberto e Mata baixa similares entre si, e mais similares com Mata fechada do que com encosta. Encosta é o ambiente menos similar aos demais.



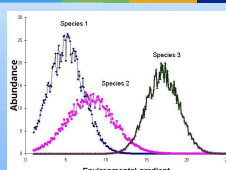
Análise de componentes principais (PCA)

- É o método mais antigo de ordenação, criado por Karl Pearson em 1901. Seu objetivo é decompor a tabela de dados com medidas correlacionadas em um novo conjunto de variáveis não correlacionadas.
- Trabalha com matriz de covariâncias, correlações ou similaridades
- Modo R
- Vantagens:
 - Redução de dimensões
 - Determinação das combinações lineares das variáveis
 - Visualização de dados multidimensionais
 - Identificação de grupos de objetos e de outliers



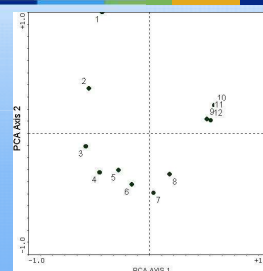
Análise de componentes principais (PCA)

- Para que os eixos (Componentes principais) sejam independentes, é necessário que a matriz tenha distribuição normal.
- O primeiro eixo deve ter a maior variância possível, enquanto o segundo eixo deve explicar a maior parte da variância residual.



Análise de componentes principais (PCA)

- Como interpretar os eixos?
- Os componentes principais devem ser interpretados como novas variáveis, as quais são independentes estatisticamente entre si. Para que essa interpretação seja feita, é necessário se identificar a contribuição de cada variável original para a formação dos eixos.



Exemplo em artigos científicos:

PUBLIC HEALTH
Spatial and Temporal Distribution of Blackflies (Diptera: Simuliidae) in the Itatiaia National Park, Brazil

RONALDO FIGUEIRO^{1,2,3}, CARLOS J.P.C. ARAÚJO-COUTINHO⁴, LEONARDO H. GIL AZEVEDO¹, ÉRIKA S. NASCIMENTO¹ E RICARDO F. MONTEIRO³

¹Depto. Entomologia, Instituto Oswaldo Cruz - Fiocruz, Av. Brasil, 4365, Manguinhos, Pavilhão Carlos Chagas 4º andar, C. postal 926, 21045-900, Rio de Janeiro, RJ

²Student of the Programa de Pós-Graduação em Ecologia of the Univ. Federal do Rio de Janeiro, RJ

³Depto. Ecologia, Instituto de Biologia, Univ. Federal do Rio de Janeiro, Av. Brigadeiro Trompowski s/n, CCS, IB, Ilha do Fundão, Cidade Universitária, C. postal, 68020, 21941-590, Rio de Janeiro, RJ

⁴Visitor researcher SUCEN / FIOCRUZ

Neotropical Entomology 35(4):542-550 (2006)

Distribuição Espacial e Temporal de Simuliídeos (Diptera: Simuliidae) no Parque Nacional de Itatiaia

Exemplo em artigos científicos:

- Objetivo do trabalho: Observar os mecanismos estruturadores de comunidades de simuliídeos em um gradiente de altitude
- Variáveis abióticas: pH, Temperatura da água, Substrato, Luminosidade, Vazão, largura do rio.
- Variáveis bióticas: *Simulium subnigrum*, *Simulium incrustatum*, *Simulium stellatum*, *Lutzimulium pernigrum*

Interpretação dos eixos e correlações

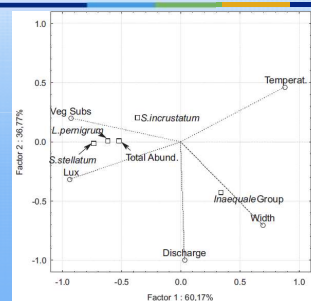
- Buscar o significado dos PCs a partir das contribuições das variáveis
- Encontrar correlações entre os PCs e as variáveis de interesse

Autumn		
	PC 1	PC 2
Vegetal substrate	0.28849	0.021612
Luminosity	0.29484	0.054124
Width	0.15965	0.26939
Discharge	0.000304	0.538549
Water temperature	0.25671	0.116325

Autumn		
	PC 1 (canopy cover)	PC 2 (river size)
<i>L. pernigrum</i>	-0.619784	0.006639
<i>S. stellatum</i>	-0.738404	-0.008418
<i>S. incrustatum</i>	-0.366747	0.209614
<i>Simulium (Inaequalum) spp.</i>	0.333463	-0.423557
Total larvae	-0.525943	0.010302

Representação Gráfica:

- 96,77% da variação explicada
- Cobertura vegetal se mostrou o fator mais importante



Estudo de caso no PAST:

- Abra o arquivo "PCA.DAT"

	X1	X2	X3	X4	X5
1	154	245	31.6	18.5	19.5
2	154	240	30.4	17.9	19.6
3	153	240	31	18.4	20.6
4	153	236	30.9	17.7	20.2
5	155	243	31.5	18.6	20.3
6	162	247	32	19	20.9
7	157	238	31.9	18.4	21.3
8	156	239	31.8	18.6	20.2
9	164	248	32.7	19.1	21.1
10	158	238	31	18.8	22
11	160	240	31.5	19.6	21

Exemplo:

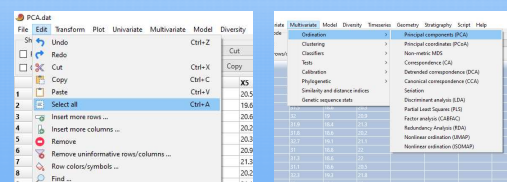
- Um pesquisador mediu cinco variáveis morfométricas em 40 andorinhas: comprimento total (X1), comprimento das asas (X2), comprimento do bico (X3), perímetro cefálico (X4) e comprimento do esterno (X5).

	X1	X2	X3	X4	X5
1	154.000	245.000	31.600	18.500	20.500
2	154.000	240.000	30.400	17.900	19.600
3	153.000	240.000	31.000	18.400	20.600
4	153.000	236.000	30.900	17.700	20.200
5	155.000	243.000	31.500	18.600	20.300
6	162.000	247.000	32.000	19.000	20.900
7	157.000	238.000	31.900	18.400	21.300
8	156.000	239.000	31.800	18.600	20.200
9	164.000	248.000	32.700	19.100	21.100
10	158.000	238.000	31.000	18.800	22.000
11	158.000	240.000	31.500	18.600	21.000
12	160.000	244.000	31.500	18.600	21.500
13	161.000	246.000	32.300	19.300	21.800
14	157.000	245.000	32.000	19.100	20.000
15	157.000	236.000	31.500	18.100	19.800
16	156.000	237.000	30.900	18.000	20.300
17	158.000	244.000	31.400	18.500	21.600



Executando a PCA:

- Vá em "edit" e selecione "select all"
- Vá no menu "multivariate" -> "Ordination" -> "Principal Components"



Interpretando os resultados:

- Em Summary: Variância explicada (eixos 1 e 2)-> $86.79+10.81=97,6\%$

Principal components analysis			
Summary	Scatter plot	Scores	Loadings
PC	Eigenvalue	% variance	
1	35.203	86.795	
2	4.38726	10.817	
3	0.609898	1.5037	
4	0.286445	0.70625	
5	0.0720632	0.17768	

Resultados:

- O que significará o PC1? Em "loadings" observemos as correlações de X1, X2, X3, X4 e X5 em relação aos componentes principais: Todos estão correlacionados positivamente, com PC1 certo?

Summary	Scatter plot	Scores	Loadings plot	Loadings	Scree plot	Spheri
	PC 1	PC 2	PC 3	PC 4	PC 5	
X1	0.53192	0.82691	-0.16823	-0.063689	0.030393	
X2	0.83194	-0.54753	-0.041131	-0.070215	-0.0382	
X3	0.096813	0.050857	0.14351	0.89196	-0.41455	
X4	0.074493	-0.012896	0.1908	0.38095	0.90153	
X5	0.10005	0.11696	0.95551	-0.22427	-0.11405	

Resultados:

- Em estudos morfométricos, é comum o eixo 1 de uma análise de componentes principais representar o tamanho do animal (Lewinsohn et al. 2002).

Summary	Scatter plot	Scores	Loadings plot	Loadings	Scree plot	Spheri
	PC 1	PC 2	PC 3	PC 4	PC 5	
X1	0.53192	0.82691	-0.16823	-0.063689	0.030393	
X2	0.83194	-0.54753	-0.041131	-0.070215	-0.0382	
X3	0.096813	0.050857	0.14351	0.89196	-0.41455	
X4	0.074493	-0.012896	0.1908	0.38095	0.90153	
X5	0.10005	0.11696	0.95551	-0.22427	-0.11405	

Considerações finais

- Devido à sua natureza complexa, estudos ambientais podem demandar técnicas capazes de simplificar a interpretação de padrões
 - A importância da abordagem multivariada diante do dilúvio de dados

