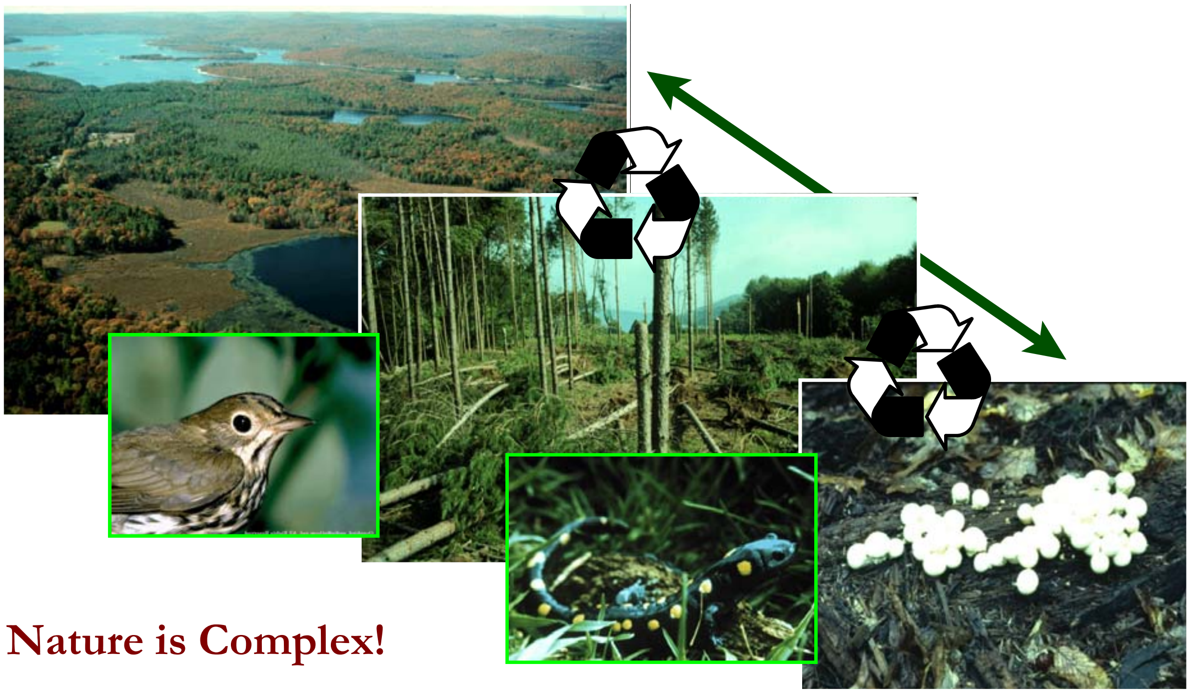
Multivariate Statistics: An Ecological Perspective



Nature is Complex!

Advantages of Multivariate Statistics

- Reflect more accurately the true multidimensional, multivariate nature of natural systems.
- Provide a way to handle large data sets with large numbers of variables.
- Provide a way of summarizing redundancy in large data sets.
- Provide rules for combining variables in an "optimal" way.

Advantages of Multivariate Statistics

- Provide a solution to the multiple comparison problem by controlling experimentwise error rate.
- Provide a means of detecting and quantifying truly multivariate patterns that arise out of the correlational structure of the variable set.
- Provide a means of exploring complex data sets for patterns and relationships from which hypotheses can be generated and subsequently tested experimentally.

What is Multivariate Statistics?

<u>**Model**</u>		<u>**Techniques**</u>
y = x1 + x2 + ... xj	→	**Regression** **Analysis of Variance** **Contingency Tables**
y1 + y2 + ... yi = x	→	**Multivariate ANOVA** **Discriminant Analysis** **CART,MRPP,MANTEL**
y1 + y2 + ... yi = x1 + x2 + ... xj	→	**Canonical Corr. Analysis** **Redundancy Analysis** **Can. Correspond. Analysis**
y1 + y2 + ... yi	→	**Ordination** **Cluster Analysis**

Multivariate Statistics (the last three rows, enclosed in a dashed box)

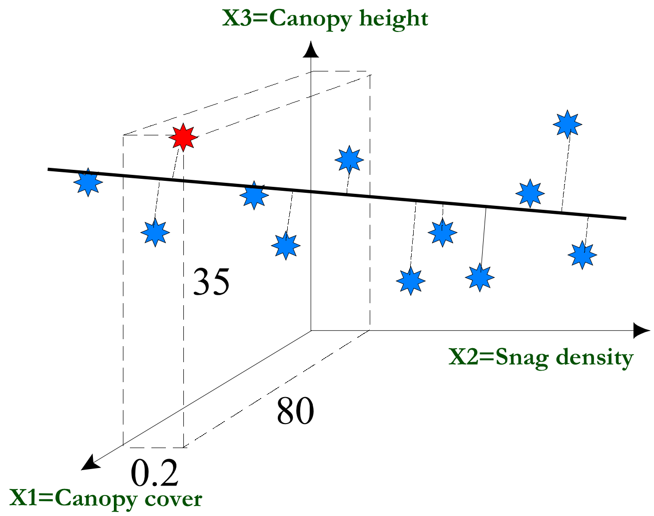
Example 1-Environmental Gradient

Data Matrix

Obs	Canopy Cover	Snag Density	Canopy Height
1	80	0.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
.	.	.	.
12	25	0.6	15

Ordination

3-Dimensional Data Space



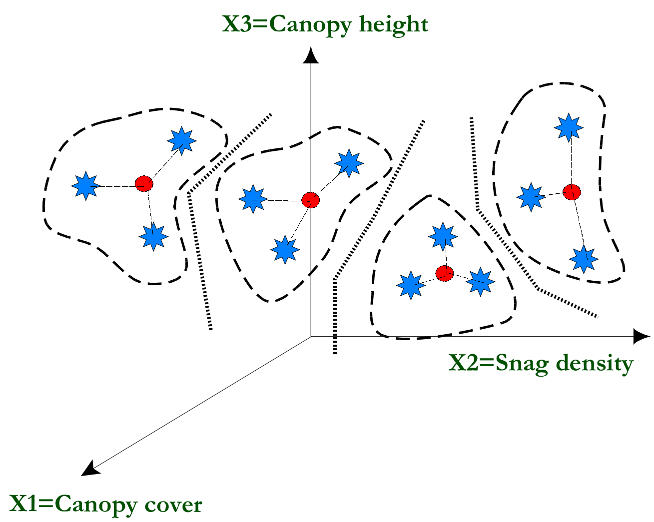
Example 1-Environmental Gradient

Data Matrix

Obs	Canopy Cover	Snag Density	Canopy Height
1	80	0.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
.	.	.	.
12	25	0.6	15

Cluster Analysis

3-Dimensional Data Space



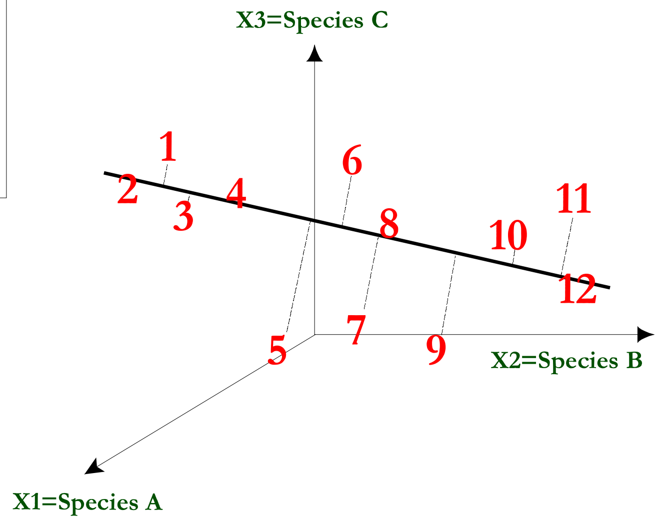
Example 2-Community Structure

Data Matrix

Sample	Species A	Species B	Species C
1	80	1.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
.	.	.	.
12	25	0.6	15

Ordination

3-Dimensional Species Space



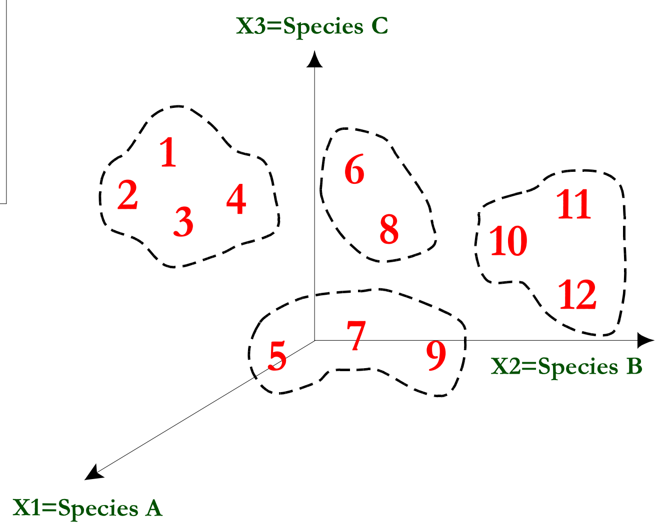
Example 2-Community Structure

Data Matrix

Sample	Species A	Species B	Species C
1	80	1.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
.	.	.	.
12	25	0.6	15

Cluster Analysis

3-Dimensional Species Space



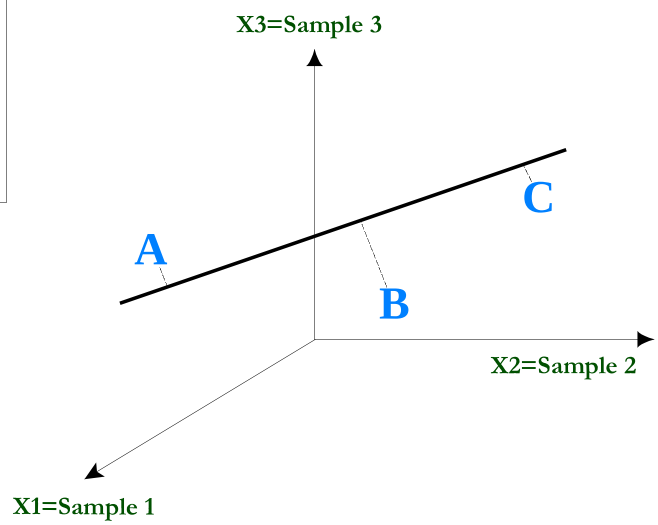
Example 2-Community Structure

Data Matrix

Sample	Species A	Species B	Species C
1	80	1.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
.	.	.	.
12	25	0.6	15

Ordination

3-Dimensional Sample Space



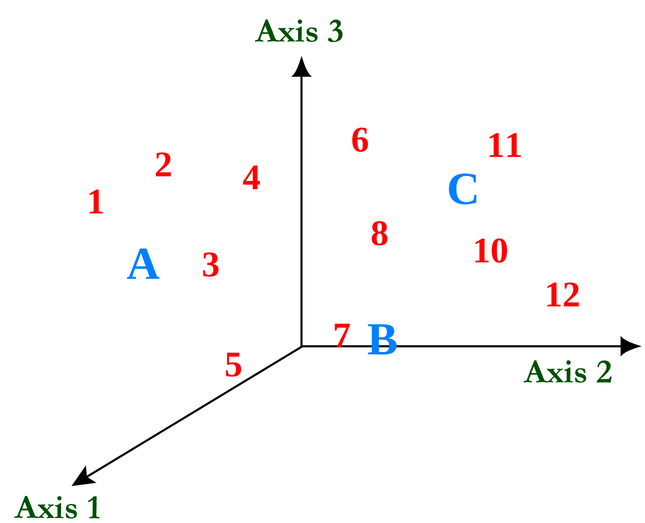
Example 2-Community Structure

Data Matrix

Sample	Species A	Species B	Species C
1	80	1.2	35
2	75	0.5	32
3	72	0.8	28
.	.	.	.
.	.	.	.
12	25	0.6	15

Ordination

3-Dimensional Ordination Space

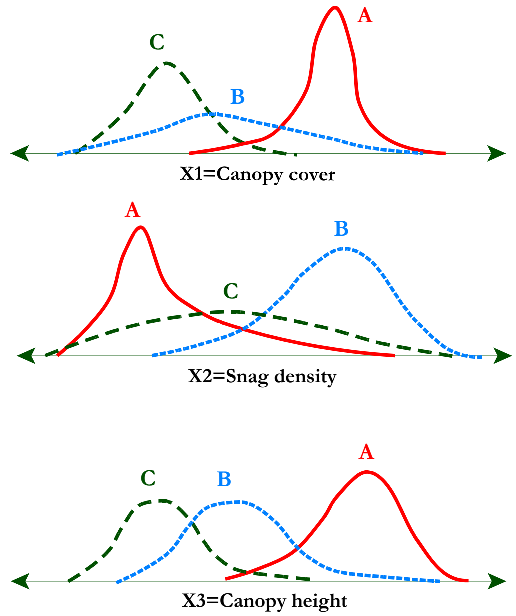


Example 3-Niche Separation

Data Matrix

Ind.	Species	Canopy Cover	Snag Density	Canopy Height
1	A	80	1.2	35
2	A	75	0.5	32
3	A	72	0.8	28
.	.	.	.	.
31	B	35	3.3	15
32	B	75	4.1	25
60	B	15	5.0	3
.	.	.	.	.
61	C	5	2.1	5
62	C	8	3.4	2
90	C	25	0.6	15

1-Dimensional Data Space

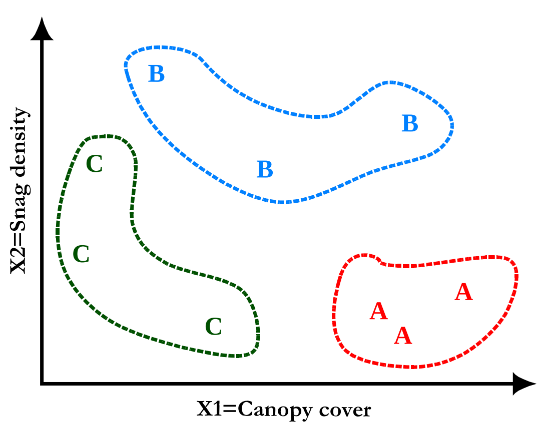


Example 3-Niche Separation

Data Matrix

Ind.	Species	Canopy Cover	Snag Density	Canopy Height
1	A	80	1.2	35
2	A	75	0.5	32
3	A	72	0.8	28
.	.	.	.	.
31	B	35	3.3	15
32	B	75	4.1	25
60	B	15	5.0	3
.	.	.	.	.
61	C	5	2.1	5
62	C	8	3.4	2
90	C	25	0.6	15

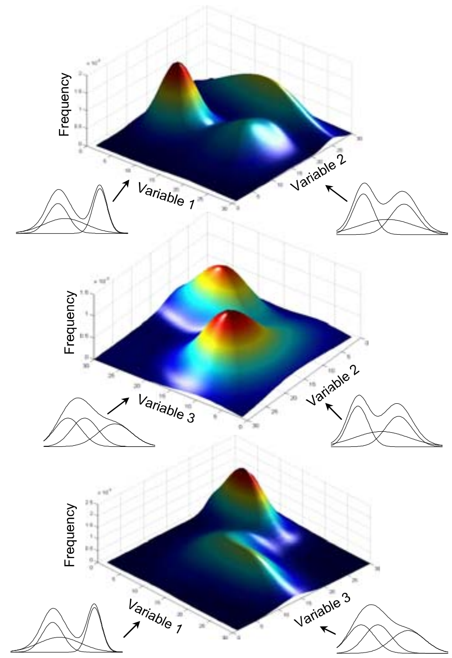
2-Dimensional Data Space



Example 3-Niche Separation

Data Matrix

Ind.	Species	Canopy Cover	Snag Density	Canopy Height
1	A	80	1.2	35
2	A	75	0.5	32
3	A	72	0.8	28
.	.	.	.	.
31	B	35	3.3	15
32	B	75	4.1	25
60	B	15	5.0	3
.	.	.	.	.
61	C	5	2.1	5
62	C	8	3.4	2
90	C	25	0.6	15

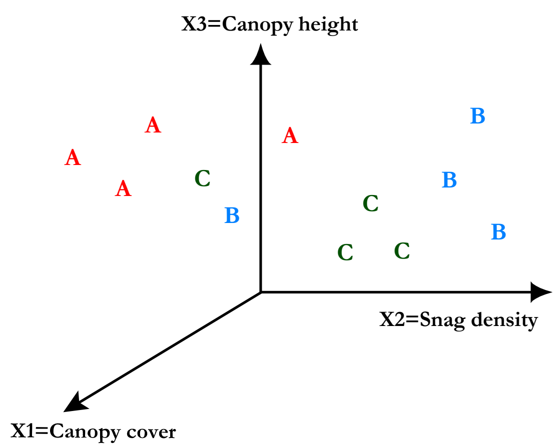


Example 3-Niche Separation

Data Matrix

Ind.	Species	Canopy Cover	Snag Density	Canopy Height
1	A	80	1.2	35
2	A	75	0.5	32
3	A	72	0.8	28
.	.	.	.	.
31	B	35	3.3	15
32	B	75	4.1	25
60	B	15	5.0	3
.	.	.	.	.
61	C	5	2.1	5
62	C	8	3.4	2
90	C	25	0.6	15

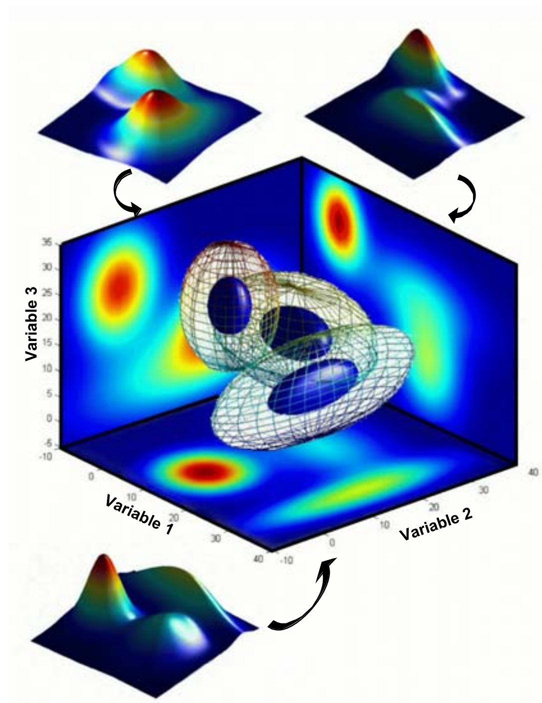
3-Dimensional Data Space



Example 3-Niche Separation

Data Matrix

Ind.	Species	Canopy Cover	Snag Density	Canopy Height
1	A	80	1.2	35
2	A	75	0.5	32
3	A	72	0.8	28
.	.	.	.	.
31	B	35	3.3	15
32	B	75	4.1	25
60	B	15	5.0	3
.	.	.	.	.
61	C	5	2.1	5
62	C	8	3.4	2
90	C	25	0.6	15

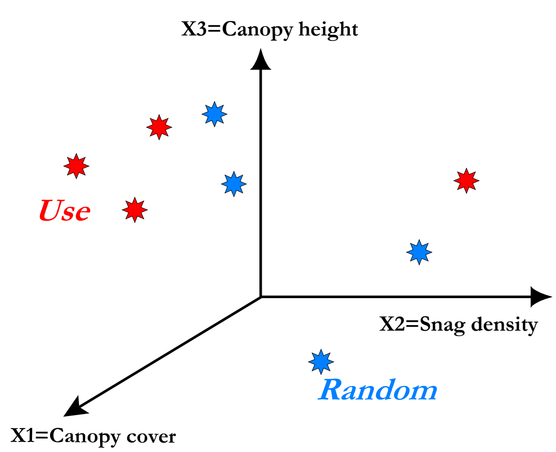


Example 4-Habitat Use

Data Matrix

Obs	Group	Canopy Cover	Snag Density	Canopy Height
1	Use	80	1.2	35
2	Use	75	0.5	32
3	Use	72	0.8	28
4	Use	35	3.3	15
.	.	.	.	.
31	Random	5	2.1	5
32	Random	68	3.4	2
33	Random	25	0.6	15
34	Random	70	1.3	33
.	.	.	.	.

3-Dimensional Data Space

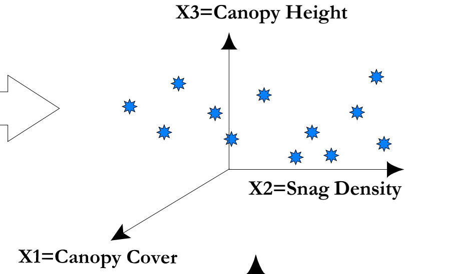


Example 5-Constrained Ordination

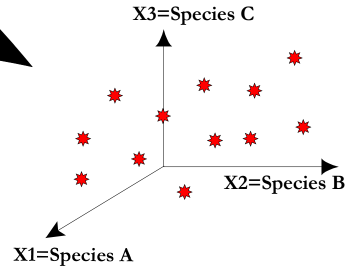
Data Matrix

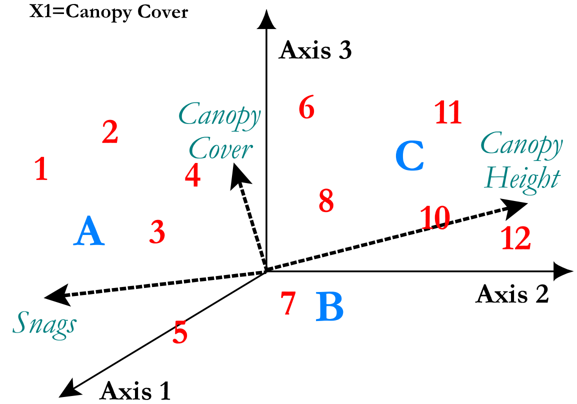
Obs	Spec A	Spec B	Spec C	C-cov	Snags	C-hgt
1	10	1	3	80	1.2	35
2	8	0	1	75	0.5	32
3	12	0	5	72	0.8	28
4	0	7	3	35	3.3	15
5	1	9	1	75	4.1	25
6	0	12	5	15	5.0	3
7	2	1	6	5	2.1	5
8	0	3	9	8	3.4	2
9	4	4	8	25	0.6	15
.	.	.	.	.	.	.

3-D Environment Space



3-D Species Space





Multivariate Statistics
Key Points

- Multivariate statistics involves cases involving multiple "dependent" variables, or a single set of variables presumed to be dependent on some underlying (latent) but unknown factors.

- All multivariate problems can be respresented as a two-way data matrix in which rows represent sampling entities and columns represent variables; the internal structure of the matrix with respect to groups of sampling entities or dependence relationships among variables distinquishes among the various multivariate techniques.

- All multivariate problems can be conceptualized geometrically as a data cloud in a *P*-dimensional data space, where the dimensions (or axes) are defined by the variables of interest; it is the shape, clumping, and dispersion of this cloud that multivariate techniques seek to describe.

Multivariate Description versus Inference

On the Descriptive Side:

- Provide rules for combining the variables in an optimal way. What is meant by ‘optimal' may vary from one technique to the next.

On the Inferential Side:

- Provide explicit control over the experimentwise error rate. Many situations in which multivariate techniques are applied could be analyzed through a series of univariate significance tests

Multivariate Confusion

- Which technique to use?
  - Ordination or cluster analysis?
  - Unconstrained or constrained ordination?
  - Polar ordination, principal components analysis, principal coordinates analysis, correspondence analysis, nonmetric multidimensional scaling?

Multivariate Confusion

- Alternative Terminology for Techniques
  - Indirect Gradient Analysis or Unconstrained Ordination?
  - Reciprocal Averaging or Correspondence Analysis?
  - Canonical Ordination or Constrained Ordination?
  - Discriminant Analysis or Canonical Variates Analysis?

21

Multivariate Confusion

- Terminology for Variable Labels Based on Data Type and Measurement Scale
  - Categorical Variable
    - Dichotomous
    - Polytomous
      - Ordinal Scale
      - Nominal Scale
  - Continuous Variable
    - Ratio Scale (true zero)
    - Interval Scale (arbitrary zero)
  - Count Variable

22

Multivariate Confusion

- Terminology for Variable Labels based on the Relationship with other Variables
  - Independent Variable
    - Variable presumed to be a cause of any change in a dependent variable; often regarded as fixed, either as in experimentation or because the context of the data suggests they play a causal role in the situation under study.
  - Dependent Variable
    - Variable presumed to be responding to a change in an independent variable; variables free to vary in response to controlled conditions.

Multivariate Techniques

Technique	Objective
Unconstrained Ordination (PCA, PO, CA, DCA, NMDS)	Extract gradients of maximum variation
Cluster Analysis (Family of techinques)	Establish groups of similar entities
Discrimination (MRPP, MANTEL, DA, CART, ...)	Test for or describe differences among groups of entities or predict group membership
Constrained Ordination (RDA, CCA, CAPS, CanCorr)	Extract gradients of variation in dependent variables explanable by independent variables

Multivariate Techniques

Technique	Dependence Type
Unconstrained Ordination (PCA, PO, CA, DCA, NMDS)	Interdependence
Cluster Analysis (Family of techinques)	Interdependence
Discrimination (MRPP, MANTEL, DA, CART, ...)	Dependence
Constrained Ordination (RDA, CCA, CAPS, CanCorr)	Dependence

Multivariate Techniques

Technique	Data Structure
Unconstrained Ordination (PCA, PO, CA, DCA, NMDS)	One set; >>2 variables
Cluster Analysis (Family of techinques)	One set; >>2 varibles
Discrimination (MRPP, MANTEL, DA, CART, ...)	Two sets; 1 grouping variable, >>2 discriminating variables
Constrained Ordination (RDA, CCA, CAPS, CanCorr)	Two sets; >>2 depend variables, >>2 independent variables

Multivariate Techniques

Obs	Group	X-set				Y-set			
1	A	a_{11}	a_{12}	a_{13}	... a_{1p}	b_{11}	b_{12}	b_{13}	... b_{1m}
2	A	a_{21}	a_{22}	a_{23}	... a_{2p}	b_{21}	b_{22}	b_{23}	... b_{2m}
3	A	a_{31}	a_{32}	a_{33}	... a_{3p}	b_{31}	b_{32}	b_{33}	... b_{3m}
.	.	.	.	.	.... .	.	.	.	.... .
.	.	.	.	.	.... .	.	.	.	.... .
n	A	a_{n1}	a_{n2}	a_{n3}	... a_{np}	b_{n1}	b_{n2}	b_{n3}	... b_{nm}
n+1	C	c_{11}	c_{12}	c_{13}	... c_{1p}				
n+2	C	c_{21}	c_{22}	c_{23}	... c_{2p}				
n+3	C	c_{31}	c_{32}	c_{33}	... c_{3p}				
.	.	.	.	.	.... .				
.	.	.	.	.	.... .				
N	C	c_{n1}	c_{n2}	c_{n3}	... c_{np}				

Multivariate Techniques

Technique	Sample Characteristics
Unconstrained Ordination (PCA, PO, CA, DCA, NMDS)	N (from known or unknown # pop's)
Cluster Analysis (Family of techinques)	N (from known or unknown # pop's)
Discrimination (MRPP, MANTEL, DA, CART, ...)	N (from known # pop's) or N1, N2, .. (from separate pop's)
Constrained Ordination (RDA, CCA, CAPS, CanCorr)	N (from one pop)

Multivariate Statistics
Key Points

- Multivariate statistics involves both <u>descriptive</u> and <u>inferential</u> statistics, although most applications are exploratory and descriptive in nature.
- Multivariate statistics includes a broad array of techniques and confusing and inconsistent use of <u>terminology</u> – *sorry, no way around this.*
- Research questions often warrant the use of more than one technique as the same technique can often be used to answer different questions and the same question can often be answered with different techniques.