Multivariate Statistics
Summary and Comparison of Techniques

- The key to multivariate statistics is understanding conceptually the relationship among techniques with regards to:
  - The kinds of problems each technique is suited for
  - The objective(s) of each technique
  - The data structure required for each technique
  - Sampling considerations for each technique
  - Underlying mathematical model, or lack thereof, of each technique
  - Potential for complementary use of techniques

Multivariate Techniques

Model	Techniques
y1 + y2 + ... + yi →	Unconstrained Ordination Cluster Analysis
$y_1 + y_2 + ... + y_i = x$ →	Multivariate ANOVA Multi-Response Permutation Analysis of Similarities Mantel Test Discriminant Analysis Logistic Regression Classification Trees Indicator Species Analysis
$y_1 + y_2 + ... + y_i = x_1 + x_2 + ... + x_j$ →	Constrained Ordination Canonical Correlation Multivariate Regression Trees

Multivariate Techniques

Technique	Objective
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS)	Extract gradients of maximum variation
Cluster Analysis (Family of techinques)	Establish groups of similar entities
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA)	Test for & describe differences among groups of entities or predict group membership
Constrained Ordination (RDA, CCA, CAP)	Extract gradients of variation in dependent variables explainable by independent variables

Multivariate Techniques

Technique	Variance Emphasis
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS)	Emphasizes variation among individual sampling entities by defining gradients of maximum total sample variance; describes the inter-entity variance structure.
Cluster Analysis (Family of techinques)	
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA)	
Constrained Ordination (RDA, CCA, CAP)	

Multivariate Techniques

Technique	Variance Emphasis
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS)	
Cluster Analysis (Family of techinques)	Emphasizes both differences and similarities among individual sampling entities by clustering entities based on inter-entity resemblance.
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA)	
Constrained Ordination (RDA, CCA, CAP)	

Multivariate Techniques

Technique	Variance Emphasis
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS)	
Cluster Analysis (Family of techinques)	
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA)	Emphasizes variation among groups of sampling entities; describes the inter-group variance structure.
Constrained Ordination (RDA, CCA, CAP)	

Multivariate Techniques

Technique	Variance Emphasis
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS)	
Cluster Analysis (Family of techinques)	
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA)	
Constrained Ordination (RDA, CCA, CAP) →	Emphasizes variation among individual sampling entities by defining gradients of maximum total sample variance explainable by environmental variables

Multivariate Techniques

Technique	Dependence Type
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS) →	Interdependence
Cluster Analysis (Family of techinques) →	Interdependence
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA) →	Dependence
Constrained Ordination (RDA, CCA, CAP) →	Dependence

Multivariate Techniques

Technique	Data Structure
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS) →	One set; >>2 variables
Cluster Analysis (Family of techinques) →	One set; >>2 varibles
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA) →	Two sets; 1 grouping variable, >>2 discriminating variables
Constrained Ordination (RDA, CCA, CAP) →	Two sets; >>2 response variables, >>2 explanatory 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}				

Unconstrained Ordination (PCA, PCO, CA, DCA, NMDS)
Cluster Analysis (Family of techinques)

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}				

Discrimination Techniques (MANOVA, MRPP, ANOSIM, Mantel; DA, LR, CART, ISA)

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}				

Constrained Ordination (RDA, CCA, CAP, COR)

Multivariate Techniques

Technique	Sample Characteristics
Unconstrained Ordination (PCA, MDS, CA, DCA, NMDS)	N (from known or unknown # pop's)
Cluster Analysis (Family of techinques)	N (from known or unknown # pop's)
Discrimination (MANOVA, MRPP, ANOSIM, Mantel, DA, LR, CART, ISA)	N (from known # pop's) or N1, N2, ... (from separate pop's)
Constrained Ordination (RDA, CCA, CAP)	N (from one pop)

Multivariate Techniques

If the research objective is to:

- Describe the major ecological gradients of variation among individual sampling entities, and/or to portray sampling entities along "continuous" gradients of maximum sample variation, then use... → Unconstrained Ordination
  - Assume linear relationship to ecological gradients... → PCA, PCO(MDS)
  - Assume unimodal relationship to ecological gradients... → CA(RA), DCA
  - Assume no particular relationship; only monotonic relationship between input and output dissimilarities... → NMDS

Multivariate Techniques

If the research objective is to:

- Establish artificial classes or groups of similar entities where pre-specified, well-defined groups do not already exist, and/or to portray sampling entities in "discrete" groups, then use... → Cluster Analysis
  - Assign entities to a specified number of groups to maximize within-group similarity or form composite clusters... → Non-hierarchical Cluster Analysis
  - Assign entities to groups and display relationships among groups as they form... → Hierarchical Cluster Analysis

Multivariate Techniques

If the research objective is to:

- Establish artificial classes or groups of entities with similar species composition and abundance where pre-specified, well-defined groups do not already exist, based on measured environmental variables, and/or to portray sampling entities in "discrete" groups representing species assemblages with distinct environmental affinities, then use... → Constrained Cluster Analysis (MRT)

Multivariate Techniques

If the research objective is to:

- Differentiate among pre-specified, well-defined classes or groups of sampling entities, and to:
  - *Test* for "significant" differences among groups...
    - Parametric test... ⇢ MANOVA / DA
    - Nonparametric tests... ⇢ MRPP, ANOSIM, Mantel

Multivariate Techniques

If the research objective is to:

- Differentiate among pre-specified, well-defined classes or groups of sampling entities, and to:
  - *Describe* the major ecological differences among groups...
    - Assume a *linear* discrimination function... ⇢ DA
    - Assume a *logistic* discrimination function... ⇢ LR (MLR)
    - Do not assume any particular function... ⇢ CART (UCT)
    - Identify "indicators" for each group... ⇢ ISA

Multivariate Techniques

If the research objective is to:

- Differentiate among pre-specified, well-defined classes or groups of sampling entities, and to:
  - *Predict* group membership of future observations...
    - *Linear* classification function... → DA (LDF)
    - *Logistic* classification function... → LR (MLR)
    - *Decision tree* classifier... → CART (UCT)
    - *Other nonparametric* classifiers... → Kernel
K nearest-neighbor

Multivariate Techniques

If the research objective is to:

- Explain the variation in a *continuous* dependent variable using two or more *continuous* independent variables, and/or to develop a model for predicting the value of the dependent variable from the values of the independent variables, then use... → Multiple Linear Regression

Alternatives: CART (URT)

Multivariate Techniques

If the research objective is to:

- Explain the variation in a *dichotomous* dependent (grouping) variable using two or more *continuous* and/or *categorical* independent variables, and/or to develop a model for predicting the group membership of a sampling entity from the values of the independent variables, then use... → Multiple Logistic Regression

  Alternatives: DA, CART (UCT)

Multivariate Techniques

If the research objective is to:

- Describe the major ecological patterns in one set of (response) variables explainable by another set of (explanatory) variables, then use... → Constrained Ordination or MRT
  - Assume *linear* response function of response variables (species) along linear gradients defined by the explanatory variables (environment)... → RDA, CAP
  - Assume *unimodal* response function of response variables (species) along linear gradients defined by the explanatory variables (environment)... → CCA, DCCA
  - Do not assume any response function... → MRT

Multivariate Techniques

If the research objective is to:

- Describe the major ecological relationships between two sets of variables expressed as distance matrices; i.e., dissimilarities between samples, then use... ⇢ Mantel Test
- Describe the major ecological relationships between two sets of variables expressed as distance matrices after accounting for a third set of variables (i.e., Y~X|Z), then use... ⇢ Partial Mantel Test

Dependence Techniques

Independent Variables

Dependent Variables		Categorical			Continuous	
		1 dicho	1 poly	>1d/p	1	>1
Categorical	1 dicho					
	1 poly					
	>1d/p					
Continuous	1					
	>1					

Dependence Techniques

CT = Contingency tables

SLR = Simple logistic regression

MLR = Multiple logistic regression

SRA = Simple linear regression

MRA = Multiple linear regression

T-test = T-test

ANOVA = Analysis of variance

UCT = Univar. classification trees

URT = Univar. regression trees

T^2-test = Hotelling's T^2

MANOVA = Multivariate analysis of variance

DA = Discriminant analysis

ISA = Indicator species analysis

RDA = Redundancy analysis

CCA = Can. correspond. analysis

CAP = Can. prin. coord. analysis

COR = Canonical corr. analysis

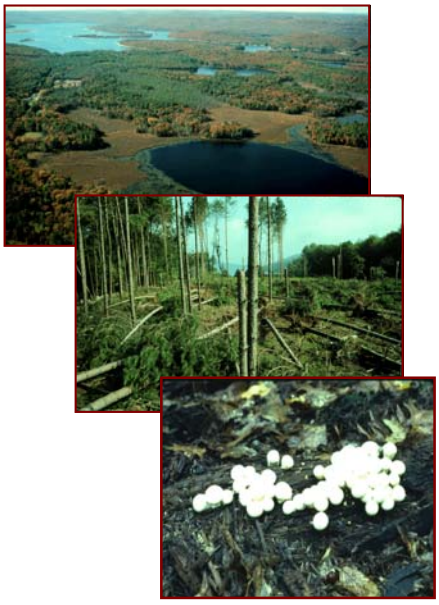
MRT = Multivar. regression trees

Dependence Techniques

Independent Variables

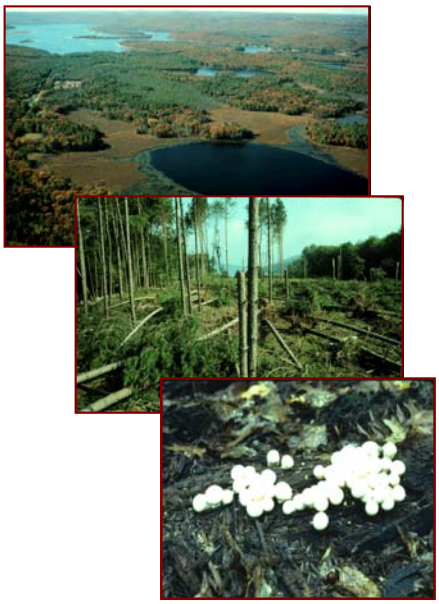
Dependent Variables		Categorical: 1 dicho	Categorical: 1 poly	Categorical: >1d/p	Continuous: 1	Continuous: >1
Categorical	1 dicho	CT SLR	CT SLR	CT MLR UCT DA	SLR	MLR UCT DA
	1 poly	CT	CT	CT UCT DA		UCT DA
	>1d/p	CT DA	CT DA	CT MRT COR RDA CAP CCA		MRT COR RDA CAP CCA
Continuous	1	SRA T-test	SRA ANOVA	MRA ANOVA URT	SRA	MRA URT
	>1	T^2-test DA ISA	Manova DA ISA	Manova MRT COR RDA CAP CCA		MRT COR RDA CAP CCA

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.