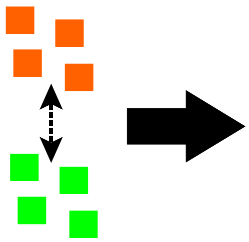
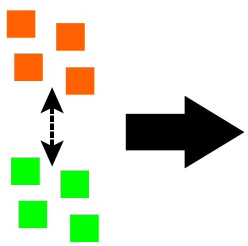
Discrimination Among Groups



- Are groups significantly different? (How valid are the groups?)
  - Multivariate Analysis of Variance [(NP)MANOVA]
  - Multi-Response Permutation Procedures [MRPP]
  - Analysis of Group Similarities [ANOSIM]
  - Mantel's Test [MANTEL]



- How do groups differ? (Which variables best distinguish among the groups?)
  - Discriminant Analysis [DA]
  - Classification and Regression Trees [CART]
  - Logistic Regression [LR]
  - Indicator Species Analysis [ISA]

Important Characteristics of Discriminant Analysis

- Essentially a *single technique* consisting of a couple of closely related procedures.
- Operates on data sets for which pre-specified, well-defined *groups already exist*.
- Assesses *dependent relationships* between one set of discriminating variables and a single grouping variable; an attempt is made to define the relationship between independent and dependent variables.

Important Characteristics of Discriminant Analysis

- *Extracts dominant, underlying gradients of variation* (canonical functions) among groups of sample entities (e.g., species, sites, observations, etc.) from a set of multivariate observations, such that variation among groups is maximized and variation within groups is minimized along the gradient.
- *Reduces the dimensionality* of a multivariate data set by condensing a large number of original variables into a smaller set of new composite dimensions (canonical functions) with a minimum loss of information.

Important Characteristics of Discriminant Analysis

- Summarizes data *redundancy* by placing similar entities in proximity in canonical space and producing a parsimonious understanding of the data in terms of a few dominant gradients of variation.
- *Describes* maximum differences among pre-specified groups of sampling entities based on a suite of discriminating characteristics (i.e., canonical analysis of discrimination).
- *Predicts* the group membership of future samples, or samples from unknown groups, based on a suite of classification characteristics (i.e., classification).

Important Characteristics of Discriminant Analysis

- Extension of *Multiple Regression Analysis* if the research situation defines the group categories as dependent upon the discriminating variables, and a single random sample (N) is drawn in which group membership is "unknown" prior to sampling.
- Extension of *Multivariate Analysis of Variance* if the values on the discriminating variables are defined as dependent upon the groups, and separate independent random samples (N1, N2, ...) of two or more distinct populations (i.e., groups) are drawn in which group membership is "known" prior to sampling.

Analogy with Regression and ANOVA

Regression Extension Analogy:

- A linear combination of measurements for two or more independent (and usually continuous) variables is used to describe or predict the behavior of a single categorical dependent variable.
- Research situation defines the group categories as dependent upon the discriminating variables.
- Samples represent a single random sample (N) of a mixture of two or more distinct populations (i.e., groups).
- A single sample is drawn in which group membership is "unknown" prior to sampling.

Analogy with Regression and ANOVA

ANOVA Extension Analogy:

- The independent variable is categorical and defines group membership (typically controlled by experimental design) and populations (i.e., groups) are compared with respect to a vector of measurements for two or more dependent (and usually continuous) variables.
- Research situation defines the discriminating variables to be dependent upon the groups.
- Samples represent separate independent random samples (N_1, N_2, ..., N_G) of two or more distinct populations (i.e., groups).
- Group membership is "known" prior to sampling and samples are drawn from each population separately.

Discriminant Analysis

Two Sides of the Same Coin

Canonical Analysis of Discriminance:

- Provides a *test* (MANOVA) of group differences and simultaneously *describes* how groups differ; that is, which variables best account for the group differences.

Classification:

- Provides a *classification* of the samples into groups, which in turn describes how well group membership can be *predicted*. The classification function can be used to predict group membership of additional samples for which group membership is unknown.

Overview of Canonical Analysis of Discriminance

- CAD seeks to *test and describe* the relationships among two or more groups of entities based on a set of two or more discriminating variables (i.e., identify boundaries among groups of entities).
- CAD involves deriving the linear combinations (i.e., *canonical functions*) of the two or more discriminating variables that will discriminate "best" among the a priori defined groups (i.e., maximize the F-ratio).
- Each sampling entity has a single composite *canonical score*, on each axis, and the group centroids indicate the most typical location of an entity from a particular group.

Hope for significant group separation and a meaningful ecological interpretation of the canonical axes.

Overview of Classification

Parametric Methods:

Valid criteria when each group is multivariate normal.

- *(Fisher's) Linear discriminant functions*: Under the assumption of equal multivariate normal distributions for all groups, derive linear discriminant functions and classify the sample into the group with the highest score. [lda(); MASS]
- *Quadratic discriminant functions*: Under the assumption of unequal multivariate normal distributions among groups, dervie quadratic discriminant functions and classify each entity into the group with the highest score. [qda(); MASS]
- *Canonical Distance*: Compute the canonical scores for each entity first, and then classify each entity into the group with the closest group mean canonical score (i.e., centroid).

Overview of Classification

<u>Nonparametric Methods:</u>

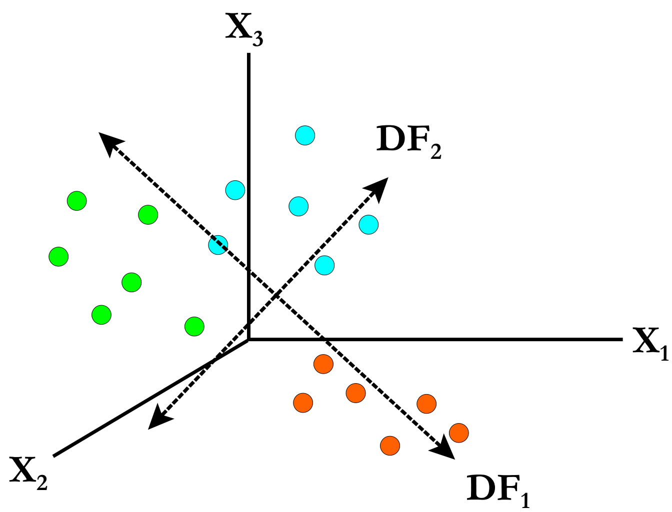
Valid criteria when no assumption about the distribution of each group can be made.

- *Kernal*: Estimate group-specific densities using a kernal of a specified form (several options), and classify each sample into the group with largest local density. [kda.kde(); ks]
- *K-Nearest Neighbor*: Classify each sample into the group with the largest local density based on user-specified number of nearest neighbors. [knn(); class]

Different classification methods will not produce the same results, particularly if parametric assumptions are not met.

Geometric View of Discriminant Analysis

- Canonical axes are derived to maximally separate the three groups on the first axis.
- The second axis is derived to provide additional separation for the blue and green groups, which overlap on the first axis.



Discriminant Analysis
The Analytical Process

- Data set
- Assumptions
- Sample size requirements
- Deriving the canonical functions
- Assessing the importance of the canonical functions
- Interpreting the canonical functions
- Validating the canonical functions

Discriminant Analysis: The Data Set

- One categorical grouping variable, and 2 or more continuous, categorical and/or count discriminating variables.
- Continuous, categorical, or count variables (preferably all continuous).
- Groups of samples must be mutually exclusive.
- No missing data allowed.
- Group sample size need not be the same; however, efficacy descreases with increasing disparity in group sizes.
- Minimum of 2 samples per group and at least 2 more samples than the number of variables.

Discriminant Analysis: The Data Set

- Common 2-way ecological data:
  - Species-by-environment
  - Species' presense/absence-by-environment
  - Behavior-by-environment
  - Sex/life stage-by-enironment/behavior
  - Soil groups-by-environment
  - Breeding demes-by-morphology
  - Etc.

			Variables			
		Group	X_1	X_2	. . .	X_p
	1	A	x_{11}	x_{12}	. . .	x_{1p}
	2	A	x_{21}	x_{22}	. . .	x_{2p}
	.	.	.	.	. . .	.
	.	.	.	.	. . .	.
	.	.	.	.	. . .	.
	n	A	x_{n1}	x_{n2}	. . .	x_{np}
Samples	n+1	B	x_{11}	x_{12}	. . .	x_{1p}
	n+2	B	x_{21}	x_{22}	. . .	x_{2p}
	.	.	.	.	. . .	.
	.	.	.	.	. . .	.
	.	.	.	.	. . .	.
	N	B	x_{N1}	x_{N2}	. . .	x_{Np}

Discriminant Analysis: The Data Set

Hammond's flycatcher: *occupied vs unoccupied sites*

?

OBS	ID	USE	GTOTAL	LTOTAL	TTOTAL	MTOTAL	OTOTAL	SNAGS6	SNAGM1	SNAGM23	SNAGM45	SNAGM6	SNAGL1	SNAGL23	SNAGL45	SNAGL6	SNAGT	BAS	BAC	BAH	BAT	FHD
1	1S0	NO	21	15	75	20	30	0	0	0	0	0	0	0	0	1	1	0	20	40	60	1.51115
2	1S1	NO	36	15	95	15	35	0	0	0	0	1	0	0	1	0	2	20	20	80	120	1.35310
3	1S2	NO	30	30	70	10	55	0	0	0	1	2	2	1	0	1	7	140	160	0	300	1.53113
4	1S3	NO	11	50	70	20	70	0	0	0	0	1	0	0	3	1	5	60	300	0	360	1.41061
5	1S4	NO	33	40	80	15	65	0	0	1	0	0	0	0	0	0	1	20	160	0	180	1.47547
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
49	1U0	YES	3	15	95	20	55	3	0	0	2	1	0	1	1	2	10	80	40	80	200	1.08919
50	1U1	YES	2	15	80	30	70	5	0	0	1	3	0	0	2	0	11	80	40	180	300	1.15219
51	1U2	YES	2	65	70	15	70	0	0	0	1	0	0	0	3	0	4	60	60	120	240	1.14216
52	1U3	YES	30	55	35	25	75	0	0	0	0	3	0	0	3	2	8	20	20	80	120	1.61978
53	1U4	YES	2	20	95	10	60	2	0	0	0	1	0	0	2	2	7	20	160	40	220	0.98561
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

DA: Assumptions

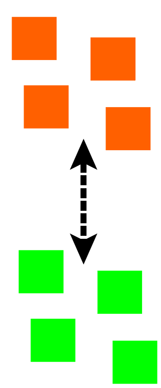
- Descriptive use of DA requires "no" assumptions!
  - However, efficacy of DA depends on how well certain assumptions are met.
- Inferential use of DA requires assumptions!
  - Evidence that certain of these assumptions can be violated moderately without large changes in correct classification results.
  - The larger the sample size, the more robust the analysis is to violations of these assumptions.

DA: Assumptions

1. Equality of Variance-Covariance Matrices:

   DA assumes that groups have equal dispersions (i.e., within-group variance-covariance structure is the same for all groups).

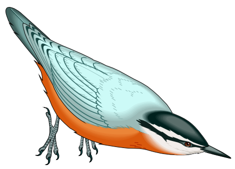
   - Variances of discriminating variables must be the same in the respective populations.
   - Correlation (or covariance) between any two variables is the same in the respective populations.



DA: Assumptions

Consequences of unequal group dispersions:

- Invalid significance tests.
- Linear canonical functions become distorted.
- Biased estimates of canonical parameters.
- Distorted representations of entities in canonical space.



- The homogeneity of covariance test can be interpreted as a significance test for habitat selectivity, and the degree of habitat specialization within a group can be inferred from the determinant of a group's covariance matrix, which is a measure of the generalized variance within the group.

DA: Assumptions

Equal group dispersions -- univariate diagnostics:

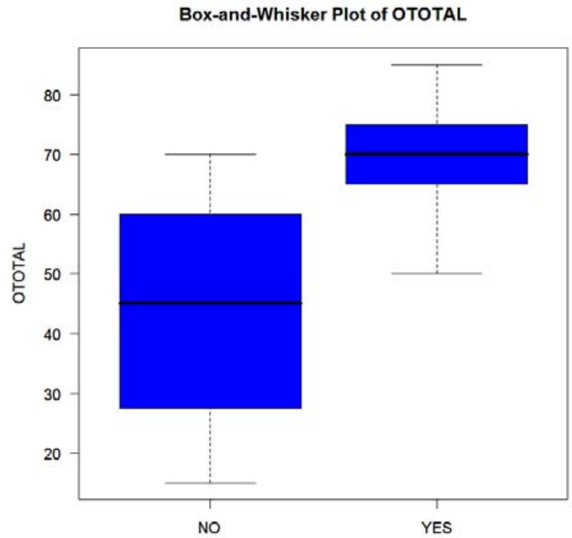
- Compute univariate test of homogeneity of variance (e.g., Fligner-Killeen nonparametric).
- Visually inspect group distributions.
  - "Univariate" homogeneity of variance does not equal "multivariate" variance-covariance homogeneity.
  - Often used to determine whether the variables should be transformed prior to the DA.
  - Usually assumed that univariate homogeneity of variance is a good step towards homogeneity of variance-covariance matrices.

DA: Assumptions

Equal group dispersions -- univariate diagnostics:

Fligner-Killeen Test of Homogeneity

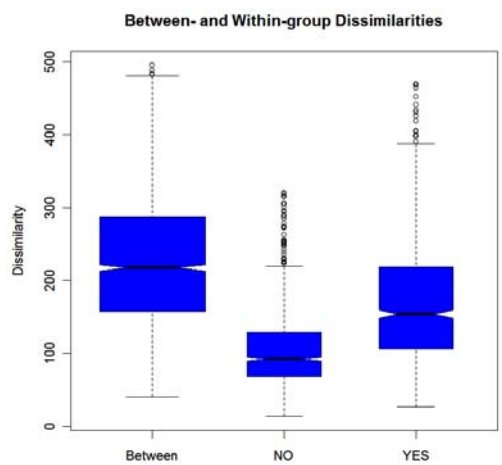
	Median chi-squared	p-value
GTOTAL	1.623	0.203
LTOTAL	12.954	0.000
TTOTAL	1.365	0.243
MTOTAL	0.217	0.641
OTOTAL	30.692	0.000
SNAGS6	18.384	0.000
SNAGM1	7.423	0.006
SNAGM23	2.588	0.108
SNAGM45	28.352	0.000
SNAGM6	0.319	0.572
SNAGL1	0.168	0.682
SNAGL23	2.190	0.139
SNAGL45	10.977	0.001
SNAGL6	10.938	0.001
BAS	12.297	0.000
BAC	13.487	0.000
BAH	0.210	0.647
FHD	23.347	0.000



DA: Assumptions

Equal group dispersions -- multivariate diagnostics:

- Conduct a multivariate test of equal group dispersions (e.g., E-test).
- Visual inspection of spread in within-group dissimilarites and canonical plots (later).

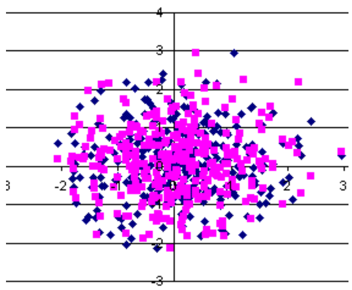


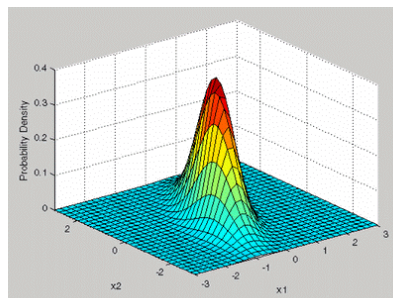

```
        Multivariate 2-sample E-test of equal distributions

data:  sample sizes 48 48, replicates 999
E-statistic = 4476.147, p-value < 2.2e-16
```

DA: Assumptions

2. Multivariate normality:

   DA assumes that the underlying structure of the data for each group is multivariate normal (i.e., hyperellipsoidal with normally varying density around the mean or centroid). Such a distribution exists when each variable has a normal distribution about fixed values on all others.





DA: Assumptions

Consequences of non-multivariate normal distributions:

- Invalid significance tests.
- Distorted posterior probabilities of group membership (i.e., will not necessarily minimize the number of misclassifications).
- In multiple CAD, second and subsequent canonical axes will not be strictly independent (i.e., orthogonal). Later canonical functions (i.e., those associated with smaller eigenvalues) will often resemble the earlier functions, but will have smaller canonical loadings.

DA: Assumptions

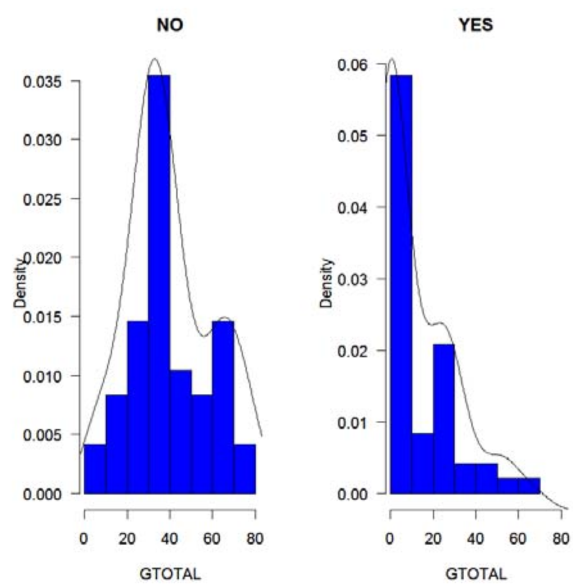
Multivariate normailty – univariate diagnostics:

- Conduct univariate tests of normality for each discriminating variable, either separately for each group or on the residuals from a one-way ANOVA with the grouping variable as the main effect).
- Visually inspect distribution plots.
  - "Univariate" normality does not equal "multivariate" normality.
  - Often used to determine whether the variables should be transformed prior to the DA.
  - Usually assumed that univariate normality is a good step towards multivariate normality.

DA: Assumptions

Multivariate normailty – univariate diagnostics:

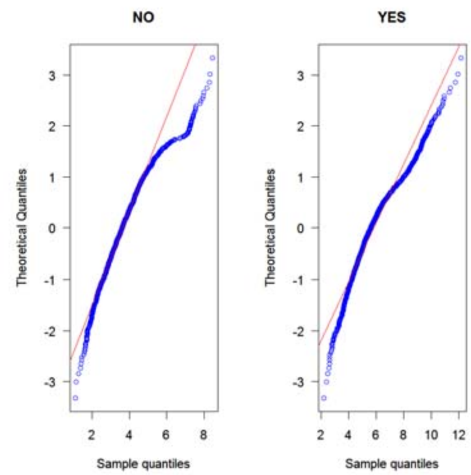
```
Anderson-Darling Test of Normality:
        Anderson-Darling A p-value
GTOTAL              3.097 < 0.001
LTOTAL              2.075 < 0.001
TTOTAL              1.821 < 0.001
MTOTAL              3.762 < 0.001
OTOTAL              0.702 0.06468
SNAGS6              2.151 < 0.001
SNAGM1             15.294 < 0.001
SNAGM23            16.043 < 0.001
SNAGM45             6.347 < 0.001
SNAGM6              2.847 < 0.001
SNAGL1             31.272 < 0.001
SNAGL23            14.995 < 0.001
SNAGL45             2.279 < 0.001
SNAGL6              4.540 < 0.001
BAS                 1.465 < 0.001
BAC                 1.172 0.00441
BAH                 6.818 < 0.001
FHD                 0.836 0.03013
```



DA: Assumptions

Multivariate normailty – multivariate diagnostics:

- Conduct a multivariate test of normality (e.g., E-statistic) separately for each group.
- Visual inspection of spread in within-group dissimilarites and canonical plots (later).



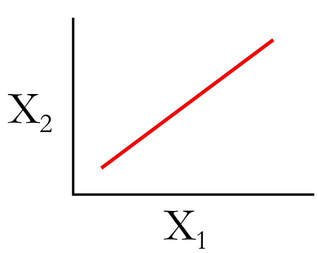

```
        Energy test of multivariate normality: estimated parameters

data:  x, sample size 96, dimension 18, replicates 999
E-statistic = 3.6248, p-value < 2.2e-16
```


DA: Assumptions

3. Singularities and multicollinearity:

DA requires that no discriminating variable be perfectly correlated with another variable (i.e., r=1) or derived from a linear combination of other variables in the data set being analyzed (i.e., the matrix must be nonsingular).

DA is adversely affected by multicollinearity, which refers to near multiple linear dependencies (i.e., high correlations) among variables in the data set.



The solution: *a priori* eliminate one or more of the offending variables.

DA: Assumptions

Consequences of multicollinearity:

- Canonical coefficients (i.e., variable weights) become difficult to interpret, because individual coefficients measure not only the influence of their corresponding original variables, but also the influence of other variables as reflected through the correlation structure.

```
Standardized Canonical Coefficients

                   CAN1

LTOTAL       1.646736324
SNAGT        0.397480978
BAH          0.650438733
GTOTAL      -0.417209741
BAS          0.313626417
SNAGL6       0.316969705
MTOTAL      -0.225091687
```


DA: Assumptions

Multicollinearity diagnostics – pairwise correlations:

- Calculate all possible pairwise correlations among the discriminating variables; high correlations (e.g., $r > 0.7$) suggest potential multicolinearity problems and indicate the need to eliminate some of the offending variables.

```
                      CORRELATION ANALYSIS

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 96

  LTOTAL    TTOTAL       BAC       BAT       FHD    OTOTAL    GTOTAL
 1.00000  -0.80786   0.77876   0.74014  -0.69086   0.64532  -0.57822
 0.0       0.0001    0.0001    0.0001    0.0001    0.0001    0.0001

   SNAGT   SNAGM45    MTOTAL       BAS    SNAGS6    SNAGM1   SNAGL45
 0.56892   0.52882  -0.49276   0.49103   0.37954   0.33786   0.26999
 0.0001    0.0001    0.0001    0.0001    0.0001    0.0008    0.0078

  SNAGL6       BAH    SNAGM6   SNAGM23   SNAGL23    SNAGL1
 0.25277  -0.22070   0.21286   0.20216   0.10905   0.01296
 0.0130    0.0307    0.0373    0.0482    0.2902    0.9003
```

DA: Assumptions

Multicollinearity diagnostics – agreement between canonical weights and loadings:

- Compare the signs and relative magnitudes of the canonical coefficients (weights) and structure coefficients (loadings) for disagreement. Pronounced differences, particularly in signs and/or rank order, indicate multicollinearity problems and highlight the need for corrective actions.

Standardized Canonical Coefficients

	CAN1
LTOTAL	1.646736324
SNAGT	0.397480978
BAH	0.650438733
GTOTAL	-0.417209741
BAS	0.313626417
SNAGL6	0.316969705
MTOTAL	-0.225091687

?

Total Canonical Structure

Variable	CAN1
LTOTAL	0.919908
SNAGT	0.762435
BAH	0.005134
GTOTAL	-0.632135
BAS	0.639319
SNAGL6	0.410062
MTOTAL	-0.452033

DA: Assumptions

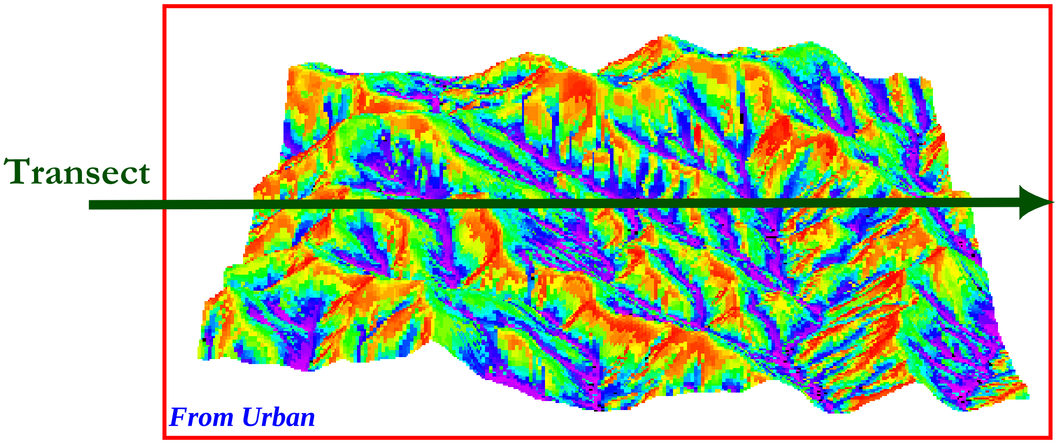
Multicollinearity solutions:

- For each pair of highly correlated variables with significant among-group differences, retain the variable with the largest F-value and/or ease of ecological interpretation, and eliminate the others.
- Use PCA to create new, completely independent composite variables from the original variables to use in DA.
- Remove one or more of the offending variables, recompute the canonical solution, and compare the results.

DA: Assumptions

4. Independent samples (& effects of outliers):

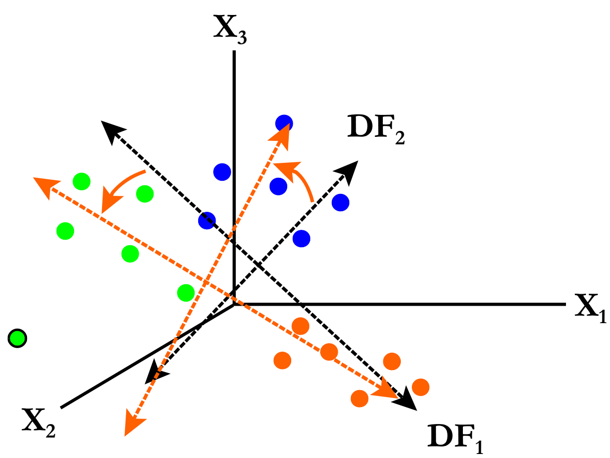
   DA assumes that random samples of observation vectors (i.e., the discriminating characteristics) have been drawn independently from respective P-dimensional multivariate normal populations.



DA: Assumptions

Consequences of non-independent samples & outliers:

- Invalid significance tests.
- Outliers exert undue pull on the direction of the canonical axes and therefore strongly affect the ecological efficacy of the analysis.



DA: Assumptions

5. Prior probabilities identifiable:

   Priors represent the probability that a sample of the i^{th} group will be submitted to the classifier; priors effect the form of the classification function.

   DA assumes that prior probabilities of group membership are *identifiable* (not necessarily equal).

   Priors may differ among groups due to unequal group population sizes, unequal sampling effort among groups, or any number of other factors.

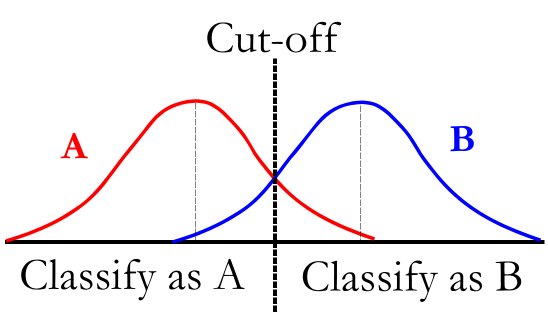
DA: Assumptions

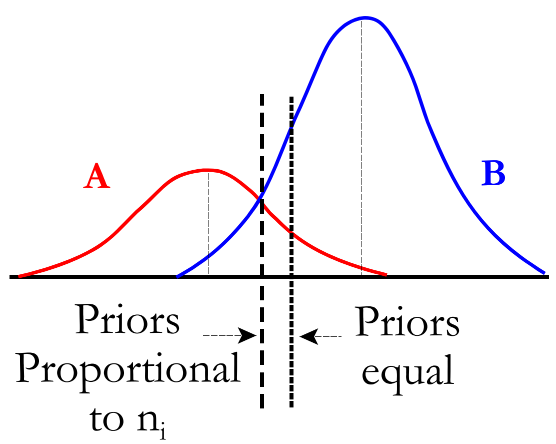
Effects of Prior Probabilities:

N	Priors	Actual	Predicted: Nest	Predicted: Not nest	#errors
7	0.5	Nest	0.83 (5.81)	0.17 (1.19)	1.19
93	0.5	Not nest	0.17 (15.81)	0.83 (77.19)	15.81
				total errors	17.00
				error rate	17.0%

Goshawk example (McCune and Grace)

Priors	Actual	Predicted: Nest	Predicted: Not nest	#errors
0.07	Nest	0.48 (3.36)	0.52 (3.64)	3.64
0.93	Not nest	0.02 (1.86)	0.98 (91.14)	1.86
			total errors	5.50
			error rate	5.5%





DA: Assumptions

Consequences of incorrect priors:

- Prior probabilities influence the forms of the classification functions. Thus, an incorrect or arbitrary specification of prior probabilities can lead to incorrect classification of samples.
- If priors are estimated by relative sampling intensities or some other estimate that actually bears no direct relationship to them, then an uncontrolled and largely inscrutable amount of arbitrariness is introduced into the DA.

DA: Assumptions

Incorrect priors diagnostics: None!

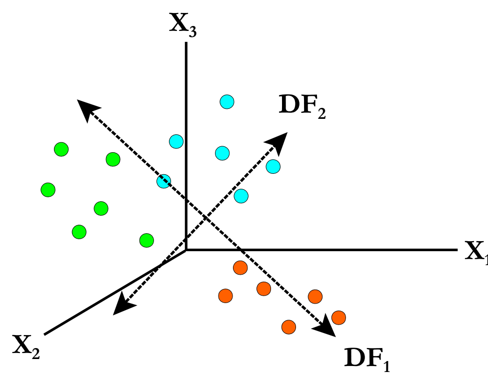
Specifying correct priors solutions:

- Use ancillary information about organisms.
- Use group sample sizes (i.e., priors proportional).
- Guess.

DA: Assumptions

6. Linearity:

The appropriateness and effectiveness DA depends on the *implicit* assumption that variables change linearly along underlying gradients and that there exits linear relationships among the variables such that they can be combined in a linear fashion to create canonical functions.

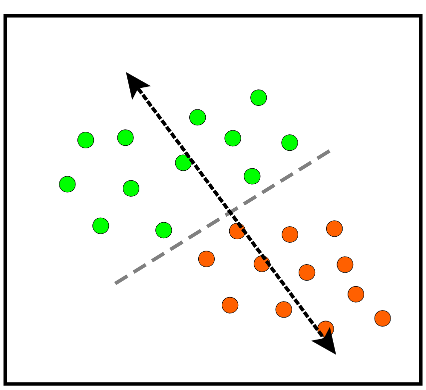


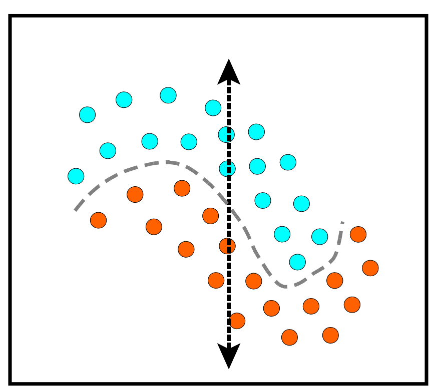
$DA1 = .8x_1 + .3x_2 -.2x_3$
$DA2 = .4x_1 - .8x_2 + .2x_3$

DA: Assumptions

Consequences of nonlinearity:

- Real nonlinear patterns will go undetected unless appropriate nonlinear transformations can be applied to model such relationships within a linear computational routine.

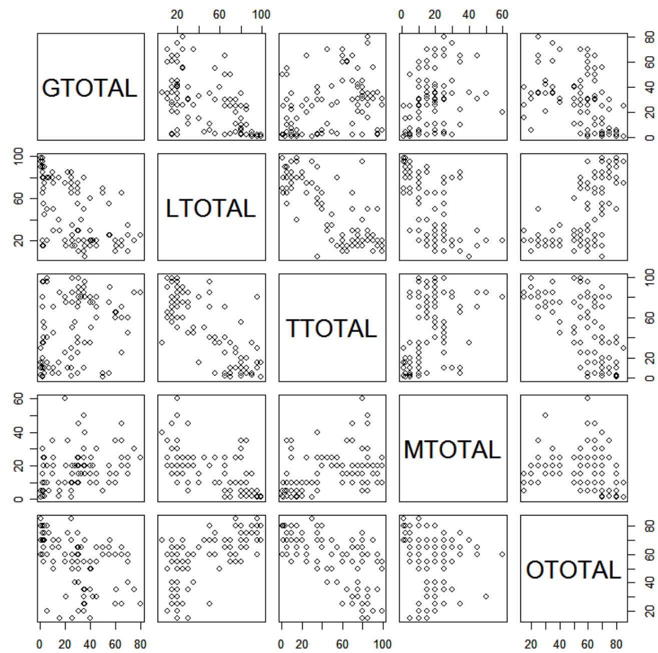




DA: Assumptions

Linearity diagnostics:

- Scatter plots of discriminating variables
- Scatter plots of canonical functions (later)



DA: Assumptions

Solutions to Violation of Assumptions:

- Calculate the canonical functions and judge their ecological significance by whether they have an ecologically meaningful and consistent interpretation.
- Evidence that procedure is moderately robust to violations of assumptions.
- Pretend there is no problem, but do not make inferences.
- Try alternative methods such as nonparametric DA (e.g., kda.kde or knn) or CART.

DA: Sample Size Considerations

General Rules:

- Minimum of at least two more samples (rows) than variables (columns).
- Minimum of at least two samples (rows) per group.
- Enough samples of each group should be taken to ensure that means and dispersions are estimated accurately and precisely.

Rule of Thumb: Each group, $n \geq (3 \cdot P)$
(Williams and Titus 1988)

DA: Sample Size Considerations

Sample Size Solutions:

- Sample sequentially until the mean and variance of the parameter estimates stabilize.
- Examine the stability of the results using a resampling procedure.
- Use variable selection procedures to reduce the number of variables.
- Divide the variables into two or more groups of related variables and conduct separate DA's on each group.
- Interpret findings cautiously.

Deriving the Canonical Functions

Selection of Variables

Reasons for using variable selection procedures:

- Data collected on many "suspected" discriminators with the specified aim of identifying the most useful.
- Data collected on many "redundant" variables with the aim of identifying a smaller subset of independent (i.e., unrelated discriminators).
- Need to reduce the number of variables to meet sample-to-variable ratio.
- Seek a parsimonious solution.

Although variable selection procedures produce an "*optimal*" set of discriminating variables, they do not guarantee the "best" (maximal) combination, and they have been heavily criticized.

Deriving the Canonical Functions

Selection of Variables

Stepwise Procedure based on Wilk's Lambda statistic:

- Forward variable selection procedure that selects the variable at each step that minimizes the overall Wilk's lambda statistic, so long as the partial Wilk's lambda is significant.
  - Likelihood ratio statistic (multivariate generalization of the F-statistic) for testing the hypothesis that group means are equal in the population.
  - Lambda approaches zero if any two groups are well separated.

Deriving the Canonical Functions

Selection of Variables

Stepwise Procedure based on Wilk's Lambda statistic:

Values calculated in each step of the selection procedure:

	vars	Wilks.lambda	F.statistics.overall	p.value.overall	F.statistics.diff	p.value.diff
1	LTOTAL	0.2859161	234.76783	2.683889e-27	234.767829	2.683889e-27
2	SNAGT	0.2146002	170.18204	8.332902e-32	30.905750	2.525220e-07
3	BAH	0.1903215	130.46421	5.000650e-33	11.736131	9.141324e-04
4	GTOTAL	0.1735867	108.30844	9.651719e-34	8.772943	3.890110e-03
5	TTOTAL	0.1599567	94.53044	2.775073e-34	7.668928	6.808555e-03
6	SNAGM45	0.1550903	80.80989	7.287388e-34	2.792658	9.817094e-02
7	BAS	0.1519241	70.17664	2.794332e-33	1.833947	1.790905e-01
8	SNAGL6	0.1481766	62.51717	8.331688e-33	2.200336	1.415534e-01
9	MTOTAL	0.1451029	56.29809	2.807526e-32	1.821701	1.806125e-01

Deriving the Canonical Functions

Eigenvalues and Associated Statistics

Characteristic Equation: $$|A - \lambda W| = 0$$

Where: A = among-groups sums-of-squares and cross products matrix
W = within-groups sums-of-squares and cross products matrix
λ = vector of eigenvalue solutions

- An NxP data set with G groups has Q (equal to G-1 or P, whichever is smaller) eigenvalues.
- Eigenvalues represent the variances of the corresponding canonical functions; they measure the extent of group differentiation along the dimension specified by the canonical function.
- $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_Q$

Deriving the Canonical Functions

Eigenvectors and Canonical Coefficients

Characteristic Equation: $$\left|A - \lambda_i W\right| v_i = 0$$

Where: λ_i = eigenvalue corresponding to the i^{th} canonical function
v_i = eigenvector associated with the i^{th} eigenvalue

- Eigenvectors are the coefficients of the variables in the linear equations that define the canonical functions and are referred to as canonical coefficients (or canonical weights).
- Uninterpretable as coefficients, and the scores they produce for entities have no intrinsic meaning, because these are weights to be applied to the variables in "raw-score" scales to produce "raw" canonical scores.

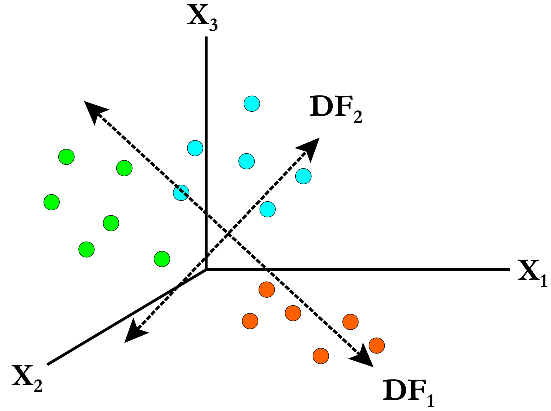
Deriving the Canonical Functions

Eigenvalues and Eigenvectors

Geometric Perspective:

- Eigenvalues equal the ratio of the between-and within-group standard deviations on the linear discriminant variables, which are defined by the eigenvectors

$DF1 = .8x_1 + .3x_2 - .2x_3$
$DF2 = .4x_1 - .8x_2 + .2x_3$



Assessing the Importance of the Canonical Functions

- How important (significant) is a canonical function?
- In multiple CAD, how many functions to retain?

1. Relative Percent Variance Criterion:

- Compare the relative magnitudes of the eigenvalues to see how much of the total discriminatory power each function accounts for.

$$\Phi_i = \frac{\lambda_i}{\sum_{i=1}^{Q} \lambda_i}$$

Assessing the Importance of the Canonical Functions

1. Relative Percent Variance Criterion:

- Measures how much of the total discriminatory power (i.e., total among-group variance) of the variables is accounted for by each canonical function.
- The cumulative percent variance of all canonical functions is equal to 100%.
- Φ_i may be very high even though group separation is minimal, because Φ does not measure the "extent" of group differentiation; it measures how much of the total differentiation is associated with each axis, regardless of the absolute magnitude in group differentiation.
- Should only be used in conjunction with other measures such as canonical correlation.

Assessing the Importance of the Canonical Functions

2. Canonical Correlation Criterion:

- Multiple correlation between the grouping variable and the corresponding canonical function (i.e., ANOVA on canonical scores).
- Ranges between zero and one; a value of zero denotes no relationship between the groups and the canonical function, while large values represent increasing degrees of association.
- *Squared canonical correlation* equals the proportion of total variation in the corresponding canonical function explained by differences in group means.

Assessing the Importance of the Canonical Functions

3. Significance Tests:

- When the data are from a *sample*, as opposed to the entire population.
- Assume independent random samples to ensure valid probability values; also multivariate normality and equal covariance matrices for parametric tests.
  - *Null Hypothesis*: The canonical correlation is equal to zero in the population.
  - *Alternative Hypothesis*: The canonical correlation is greater than zero in the population.

Assessing the Importance of the Canonical Functions

3. Significance Tests:

Cautions!

- Function may not discriminate among the groups well enough (i.e., a small canonical correlation).
- Function may fail to correctly classify enough entities into their proper groups (i.e., a poor correct classification rate).
- Function may not have a meaningful ecological interpretation as judged by the canonical loadings.
- Ultimately, the utility of each canonical function must be grounded on ecological criteria.

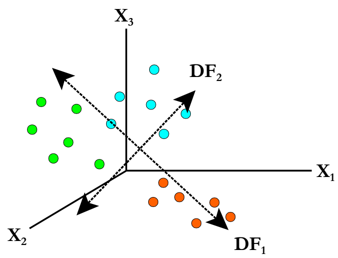
Assessing the Importance of the Canonical Functions

Canonical correlation and significance tests:

```
Call:
lm(formula = y.lda.pred$x ~ grp)

Residuals:
     Min       1Q   Median       3Q      Max 
-3.05349 -0.78300  0.00691  0.61421  2.12551 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -2.3002     0.1443  -15.94   <2e-16 ***
grpYES        4.6005     0.2041   22.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 1 on 94 degrees of freedom
Multiple R-Squared: 0.8438,     Adjusted R-squared: 0.8422 
F-statistic:   508 on 1 and 94 DF,  p-value: < 2.2e-16
```

Assessing the Importance of the Canonical Functions

4. Canonical Scores & Associated Plots:



$$z_{ij} = c_{i1}x^*_{j1} + c_{i2}x^*_{j2} + \ldots + c_{iP}x^*_{jP}$$

Z_{ij} = standardized canonical score for i^{th} canonical function and j^{th} sample

c_{ik} = standardized canonical coefficient for i^{th} function and k^{th} variable

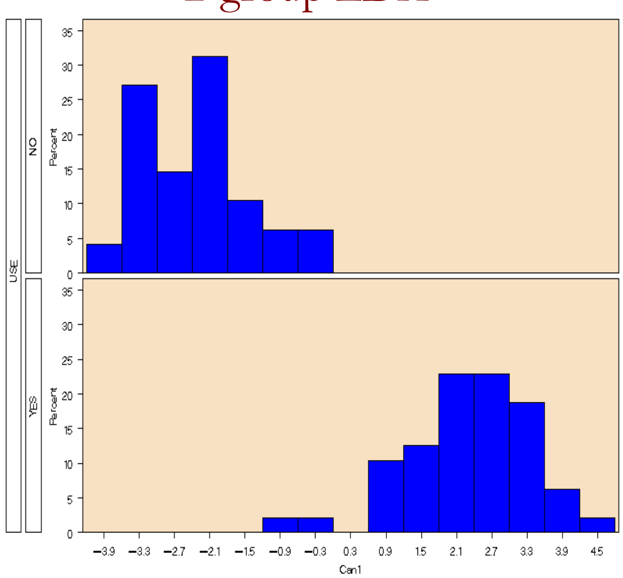
x^*_{jk} = standardized value for j^{th} sample and k^{th} variable

- Gaphically illustrate the relationships among entities, since entities in close proximity in canonical space are ecologically similar with respect to the environmental gradients defined by the canonical functions.
- Typically used to assess how much overlap exists in group distributions; i.e., how distinct the groups are.

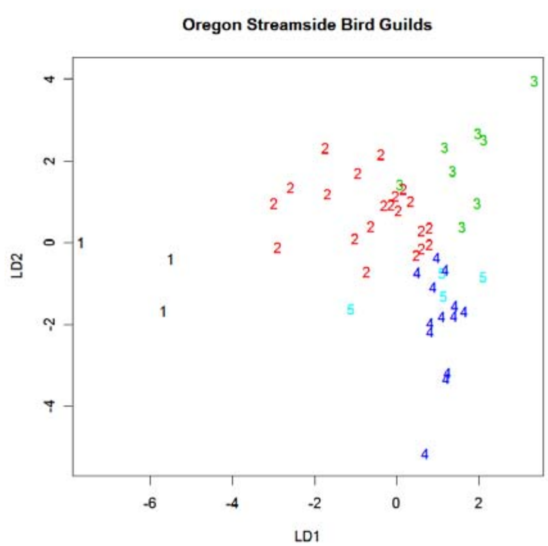
Assessing the Importance of the Canonical Functions

4. Canonical Scores & Associated Plots:

2-group LDA



Mulit-group LDA



Assessing the Importance of the Canonical Functions

5. Classification Accuracy:

- Measure the accuracy of the classification criterion to indirectly assess the amount of canonical discrimination contained in the variables. The higher the correct classification rate, the greater the degree of group discrimination achieved by the canonical functions.
  - *Classification (or confusion) matrix* provides the number and percent of sample entities classified correctly or incorrectly into each group.
  - *Correct classification rate* is the percentage of samples classified correctly.

Assessing the Importance of the Canonical Functions

5. Classification Accuracy:

Quadratic classification criterion (Mahalanobis distance)

```
Resubstitution Results using Quadratic Discriminant Function

             Generalized Squared Distance Function
             2            _        -1    _
            D (X) = (X-X )' COV   (X-X ) + ln |COV |
             j           j       j     j            j

        Posterior Probability of Membership in Each USE
                            2                    2
         Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D (X))
                            j           k            k

          Posterior Probability of Membership in USE
               From       Classified
        Obs    USE        into USE            NO        YES

          1    NO         NO              0.9999     0.0001
          2    NO         NO              0.9999     0.0001
          3    NO         NO              0.8767     0.1233
          4    NO         NO              0.8445     0.1555
          5    NO         NO              0.9999     0.0001
          6    NO         NO              0.9995     0.0005
          7    NO         NO              0.9999     0.0001
          8    NO         NO              1.0000     0.0000
          9    YES        YES             0.0508     0.9492
         10    YES        YES             0.0000     1.0000
         11    YES        YES             0.0000     1.0000
         12    YES        YES             0.0123     0.9877
         13    YES        NO         *    0.8263     0.1737
         14    YES        YES             0.0000     1.0000
         15    YES        YES             0.0000     1.0000
```

Classification Matrix

```
Number of Observations and Percent Classified into USE
    From
    USE              NO            YES          Total

    NO               48              0             48
                 100.00           0.00         100.00


    YES               1             47             48
                   2.08          97.92         100.00


    Total            49             47             96
                  51.04          48.96         100.00


    Priors          0.5            0.5


            Error Count Estimates for USE
                          NO            YES          Total

   Rate               0.0000         0.0208         0.0104
   Priors             0.5000         0.5000
```

Missclassified

Assessing the Importance of the Canonical Functions

5. Classification Accuracy:

Classification Matrix

Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into USE

From USE	NO	YES	Total
NO	48	0	48
	100.00	0.00	100.00
YES	1	47	48
	2.08	97.92	100.00
Total	49	47	96
	51.04	48.96	100.00
Priors	0.5	0.5	

Error Count Estimates for USE

	NO	YES	Total
Rate	0.0000	0.0208	0.0104
Priors	0.5000	0.5000	

Jacknife Cross-Validation Classification Matrix

Cross-validation Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into USE

From USE	NO	YES	Total
NO	44	4	48
	91.67	8.33	100.00
YES	1	47	48
	2.08	97.92	100.00
Total	45	51	96
	46.88	53.13	100.00
Priors	0.5	0.5	

Error Count Estimates for USE

	NO	YES	Total
Rate	0.0833	0.0208	0.0521
Priors	0.5000	0.5000	

Assessing the Importance of the Canonical Functions

5. Classification Accuracy – chance-corrected

- A certain percentage of samples in any data set are expected to be correctly classified by chance, regardless of the classification criterion.
  - Expected probability of classification into any group by chance is proportional to the group size.
  - As the relative size of any single group becomes predominant, the correct classification rate based on chance alone tends to increase towards unity.
  - The need for a "chance-corrected" measure of prediction (or discrimination) becomes greater with more dissimilar group sizes (or prior probabilities).

Assessing the Importance of the Canonical Functions

5. Classification Accuracy – chance-corrected

(A) Maximum Chance Criterion (C_{max})

- Appropriate when prior probabilities are assumed to be equal to group sample sizes.
- Should be used only when the sole objective is to maximize the "overall" correct classification rate.

	No	Yes	Total
No	48	0	48
Yes	1	47	48
Total	49	47	96
Priors	*.50*	*.50*	

$C_{max} = .5$ $C_{obs} = .99$

	No	Yes	Total
No	20	0	20
Yes	1	75	76
Total	21	75	96
Priors	*.21*	*.79*	

$C_{max} = .79$ $C_{obs} = .99$

Assessing the Importance of the Canonical Functions

5. Classification Accuracy – chance-corrected

(B) Proportional Chance Criterion (C_{pro})

- Appropriate when prior probabilities are assumed to be equal to group sample sizes.
- Use only when objective is to maximize the "overall" correct classification rate.

$$C_{pro} = p^2 + (1 - p)^2$$

P = proportion of samples in group 1

1-P = proportion of samples in group 2

	No	Yes	Total
No	48	0	48
Yes	1	47	48
Total	49	47	96
Priors	*.50*	*.50*	

$C_{pro} = .5$ $C_{obs} = .99$

	No	Yes	Total
No	20	0	20
Yes	1	75	76
Total	21	75	96
Priors	*.21*	*.79*	

$C_{pro} = .67$ $C_{obs} = .99$

Assessing the Importance of the Canonical Functions

5. Classification Accuracy – chance-corrected

(C) Tau Statistic

- Appropriate when prior probabilities are known or are not assumed to be equal to sample sizes.
- Tau = percent reduction in errors over random assignment.

$$Tau = \frac{n_o - \sum_{i=1}^{G} p_i n_i}{n - \sum_{i=1}^{G} p_i n_i}$$

$n =$ total # samples
$n_o =$ # samples correctly classified
$n_i =$ # samples in i^{th} group
$p_i =$ prior probability in the i^{th} group

	No	Yes	Total
No	48	0	48
Yes	1	47	48
Total	49	47	96
Priors	*.50*	*.50*	

Tau = .98 C_{obs} = .99

	No	Yes	Total
No	20	0	20
Yes	1	75	76
Total	21	75	96
Priors	*.21*	*.79*	

Tau = .97 C_{obs} = .99

Assessing the Importance of the Canonical Functions

5. Classification Accuracy – chance-corrected

(D) Kappa Statistic

- Appropriate when prior probabilities are assumed to be equal to sample sizes.
- Kappa = percent reduction in errors over random assignment.

$$Kappa = \frac{p_o - \sum_{i=1}^{G} p_i q_i}{1 - \sum_{i=1}^{G} p_i q_i}$$

$p_o =$ % samples correctly classified
$p_i =$ % samples in i^{th} group
$q_i =$ % samples classified into i^{th} group

	No	Yes	Total
No	48	0	48
Yes	1	47	48
Total	49	47	96
Priors	*.50*	*.50*	

Kappa = .98 C_{obs} = .99

	No	Yes	Total
No	20	0	20
Yes	1	75	76
Total	21	75	96
Priors	*.21*	*.79*	

Kappa = .97 C_{obs} = .99

Assessing the Importance of the Canonical Functions

5. Classification Accuracy – chance-corrected

All four criteria are unbiased only when computed with "holdout" samples (i.e., split-sample approach).

Classification Summary for Calibration Data: $\boxed{\text{DABOOK.TRAINING}}$

Number of Observations and Percent Classified into USE

From USE	NO	YES	Total
NO	24 100.00	0 0.00	24 100.00
YES	1 4.17	23 95.83	24 100.00
Total	25 52.08	23 47.92	48 100.00
Priors	0.5	0.5	

Error Count Estimates for USE

	NO	YES	Total
Rate	0.0000	0.0417	$\boxed{0.0208}$
Priors	0.5000	0.5000	

Kappa = .96 C_{obs} = .98

Classification Summary for Test Data: $\boxed{\text{DABOOK.VALIDATE}}$

Number of Observations and Percent Classified into USE

From USE	NO	YES	Total
NO	23 95.83	1 4.17	24 100.00
YES	2 8.33	22 91.67	24 100.00
Total	25 52.08	23 47.92	48 100.00
Priors	0.5	0.5	

Error Count Estimates for USE

	NO	YES	Total
Rate	0.0417	0.0833	$\boxed{0.0625}$
Priors	0.5000	0.5000	

Kappa = .88 C_{obs} = .94

Interpreting the Canonical Functions

1. Standardized Canonical Coefficients (Canonical Weights):

$$c_i = u_i \sqrt{\frac{w_{ii}}{n-g}}$$

u_i = vector of raw canonical coefficients associated with the i^{th} eigenvalue

w_{ii} = sums-of-squares for the i^{th} variable, or the i^{th} diagonal element of the within-groups sums-of-squares and cross products matrix

n = number of samples

g = number of groups

- Weights that would be applied to the variables in "standardized" form to generate "standardized" canonical scores.
- Measure the "relative" contribution of the variables.

Standardized canonical coefficients may distort the true relationship among variables in the canonical functions when the correlation structure of the data is complex.

Interpreting the Canonical Functions

2. Total Structure Coefficients (Canonical Loadings):

$$s_{ij} = \sum_{k=1}^{p} r_{jk} c_{ik}$$

r_{jk} = total correlation between the j^{th} and k^{th} variables
c_{ik} = standardized canonical coefficient for the i^{th} canonical function and k^{th} variable

- Bivariate product-moment correlations between the canonical functions and the individual variables.
- Structure coefficients generally are not affected by relationships with other variables.
- We can define a canonical function on the basis of the structure coefficients by noting the variables that have the largest loadings.
- The squared loadings indicate the percent of the variable's variance accounted for by that function.

Interpreting the Canonical Functions

Canonical Coefficients & Loadings:

Total-Sample Standardized Canonical Coefficients

Variable	Label	Can1
LTOTAL	LTOTAL	1.646736324
SNAGT	SNAGT	0.397480978
BAH	BAH	0.650438733
GTOTAL	GTOTAL	-0.417209741
BAS	BAS	0.313626417
SNAGL6	SNAGL6	0.316969705
MTOTAL	MTOTAL	-0.225091687

Total Canonical Structure

Variable	Label	Can1
LTOTAL	LTOTAL	0.919908
SNAGT	SNAGT	0.762435
BAH	BAH	0.005134
GTOTAL	GTOTAL	-0.632135
BAS	BAS	0.639319
SNAGL6	SNAGL6	0.410062
MTOTAL	MTOTAL	-0.452033

No ⟵⟶ Yes

No: GTOTAL, MTOTAL

Yes: LTOTAL, SNAGT, BAS, SNAGL6

Interpreting the Canonical Functions

3. Covariance-Controlled Partial F-Ratios:

Partial F-ratio for each variable in the model -- the statistical significance of each variable's contribution to the discriminant model, given the relationships that exist among all of the discriminating variables.

- The relative importance of the variables can be determined by examining the absolute sizes of the significant partial F-ratios and ranking them.
- Unlike the standardized canonical coefficients and structure coefficients, the partial F is an "*aggregative*" measure in that it summarizes information across the different canonical functions. Thus, it does not allow you to evaluate each canonical function independently.

Interpreting the Canonical Functions

3. Covariance-Controlled Partial F-Ratios:

Stepwise Selection: Step 8

Statistics for Removal, DF = 1, 88

	Variable	Label	Partial R-Square	F Value	Pr > F
1	LTOTAL	LTOTAL	0.4997	87.89	<.0001
7	MTOTAL	MTOTAL	0.0260	2.35	0.1288
4	SNAGL6	SNAGL6	0.0506	4.69	0.0331
6	BAS	BAS	0.0429	3.94	0.0503
2	BAH	BAH	0.2020	22.27	<.0001
3	GTOTAL	GTOTAL	0.0866	8.34	0.0049
5	SNAGT	SNAGT	0.0443	4.08	0.0465

No variables can be removed.

Statistics for Entry, DF = 1, 87

Variable	Label	Partial R-Square	F Value	Pr > F	Tolerance
OTOTAL	OTOTAL	0.0200	1.77	0.1867	0.3379
SNAGM1	SNAGM1	0.0026	0.22	0.6377	0.3061
SNAGM23	SNAGM23	0.0103	0.91	0.3430	0.3022
SNAGL1	SNAGL1	0.0203	1.81	0.1826	0.3434
SNAGL23	SNAGL23	0.0022	0.20	0.6597	0.3434
SNAGL45	SNAGL45	0.0045	0.39	0.5334	0.3022
SNAGS6	SNAGS6	0.0165	1.46	0.2299	0.1939
SNAGM6	SNAGM6	0.0019	0.17	0.6847	0.2872

No variables can be entered.
No further steps are possible.

Interpreting the Canonical Functions

4. Potency Index:

$$PI_j = \sum_{i=1}^{M}\left[s_{ij}^2 \left(\frac{\lambda_i}{\sum_{i=1}^{M} \lambda_i} \right) \right]$$

m = number of significant or retained canonical functions
s_{ij} = structure coefficient for the i^{th} canonical function and j^{th} variable
λ_i = eigenvalue corresponding to the i^{th} canonical function

- % of the total discriminating power of the *retained* canonical functions accounted for by each variable.
- Analogous to *final communality* estimates in principal components.
- Potency index is an "*aggregative*" measure because it summarizes information across the different canonical functions.

Validating the Canonical Functions

- The results of DA are reliable only if means and dispersions are estimated accurately and precisely, particularly when the objective is to develop classification functions for predicting group membership of future observations.
- The best assurance of reliable results is an intelligent sampling plan and a large sample.
- Validation becomes increasing important as the sample size decreases relative to dimensionality (number of variables).
- Validation is particularly important when you are concerned with the external validity of the findings.

Validating the Canonical Functions

1. Split-Sample Validation (Cross-Validation):

The premise is that an upward bias will occur in the predication accuracy of the classification criterion if the samples used in deriving the classification matrix are the same as those used in deriving the classification function.

- Randomly divide the data set into two subsets;
- Compute the classification criterion using the "training", "analysis", or "calibration" data subset;
- Use the derived criterion to classify samples from the "validation", "holdout", or "test" data subset; and,
- Use the resulting correct classification rate to judge the reliability and robustness of the classification criterion.

Validating the Canonical Functions

1. Split-Sample Validation (Cross-Validation):

Classification Summary for Calibration Data: $\boxed{\text{DABOOK.TRAINING}}$

Number of Observations and Percent Classified into USE

From USE	NO	YES	Total
NO	24 100.00	0 0.00	24 100.00
YES	1 4.17	23 95.83	24 100.00
Total	25 52.08	23 47.92	48 100.00
Priors	0.5	0.5	

Error Count Estimates for USE

	NO	YES	Total
Rate	0.0000	0.0417	$\boxed{0.0208}$
Priors	0.5000	0.5000	

Kappa = .96 C_{obs} = .98

Classification Summary for Test Data: $\boxed{\text{DABOOK.VALIDATE}}$

Number of Observations and Percent Classified into USE

From USE	NO	YES	Total
NO	23 95.83	1 4.17	24 100.00
YES	2 8.33	22 91.67	24 100.00
Total	25 52.08	23 47.92	48 100.00
Priors	0.5	0.5	

Error Count Estimates for USE

	NO	YES	Total
Rate	0.0417	0.0833	$\boxed{0.0625}$
Priors	0.5000	0.5000	

Kappa = .88 C_{obs} = .94

- Poor classification rates indicate an unstable classification criterion, and that a larger sample may be required to obtain accurate and precise estimates of means and dispersions.

Validating the Canonical Functions

1. Split-Sample Validation (Cross-Validation):

   Strategies:

   - Random split into two equal sized subsets.
   - More entities in the training sample.
   - When selecting entities for the training and validation samples, a proportionately stratified random sampling procedure based on group sizes is usually employed.
   - Use a v-fold cross-validation procedure to classify each of the v subsets of the data based on v-1 subsets used to build the classification criterion; combine the results across v prediction sets.

Limitations of Discriminant Analysis

- DA is sensitive to the presence of outliers.
- When several canonical functions exist, as in multiple CAD, by only interpreting the first one or two canonical functions you may overlook a later axis that accounts for most of the discriminating power in some variable. Consequently, even though this variable has significant univariate discriminating power, this power is lost in the canonical transformation.
- DA is only capable of detecting gradients that are intrinsic to the data set. There may exist other more important discriminating gradients not measured using the selected variables, and these dominant, undetected gradients may distort or confuse any relationships intrinsic to the data.

Limitations of Discriminant Analysis

- There are different philosophies on how much weight to give to the objective measures of discriminant performance.
  - Canonical functions should be evaluated solely on whether they offer a *meaningful ecological interpretation*; little emphasis is placed on the statistics associated with the eigenvalues
  - Canonical functions should be evaluated largely on the basis of *objective performance criteria*; otherwise we can not discern whether the patterns revealed by the analysis are "real" or merely reflect sampling bias.

Limitations of Discriminant Analysis

- As with a regression equation, the canonical and classification function(s) should be *validated* by testing their efficacy with a fresh sample of entities. The observed accuracy of prediction on the sample upon which the function was developed will always be spuriously high. The true discriminatory power of the function will be found only when it is tested with a completely separate sample.

Limitations of Discriminant Analysis

- Parametric assumptions (multivariate normality, equality of covariance matrices) and linearity assumption are particularly restrictive and reduce the effectiveness of DA when the group data structure is complex.
  - Other procedures (e.g., CART) may perform better under these conditions.