Linear Discriminant Function

- LDF & MANOVA
- LDF & Multiple Regression
- Geometric example of LDF & multivariate power
- Evaluating & reporting LDF results
- 3 kinds of “weights”
- Evaluating & reporting MANOVA results

ldf & MANOVA

1 grouping variable and multiple “others” (quantitative or binary)

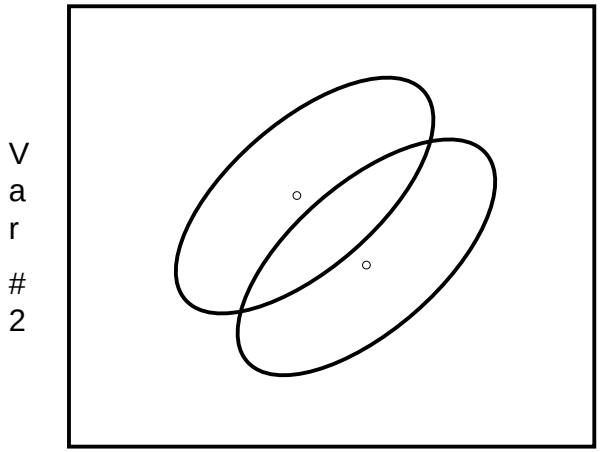
Naming conventions :

- LDF -- if the groups are “naturally occurring”
  - bio-taxonomy to diagnostic categories & measurement
  - grouping variable is called the “criterion”
  - others called the “discriminator” or “predictor” variables
- MANOVA -- if the groups are the “result of IV manipulation”
  - multivariate assessment of agricultural “programs”
  - grouping variable is called the “IV”
  - others called the “DVs”

Ways of thinking about the “new variable” in ldf/MANOVA

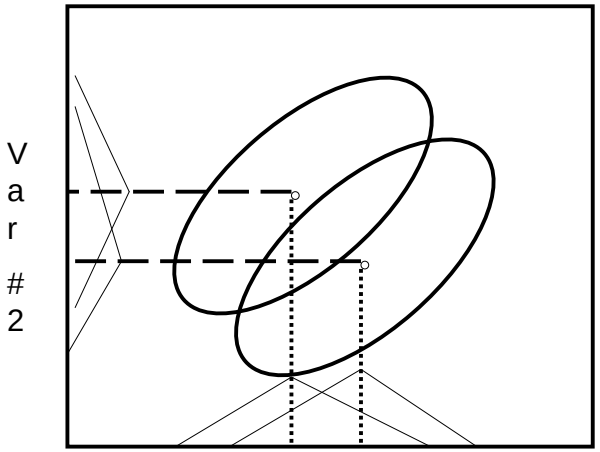
- (like regression) involves constructing a “new” quantitative variate from a weighted combination of quantitative, binary, or coded predictors, discriminators or DVs
- The “new” variable is constructed so that when it is used as the DV in an ANOVA, the F-value will be as large as possible (simultaneously maximizing between groups variation and minimizing within-groups variation)
- the “new” variable is called
  - linear discriminant function -- a linear function of the original variables constructed to maximally discriminate among the “groups”
  - MANOVA variate -- a “variate” is constructed from variables
  - canonical variate -- alludes to canonical correlation as the general model within which all corr and ANOVA models fi

How ldf works -- two groups and 2 vars



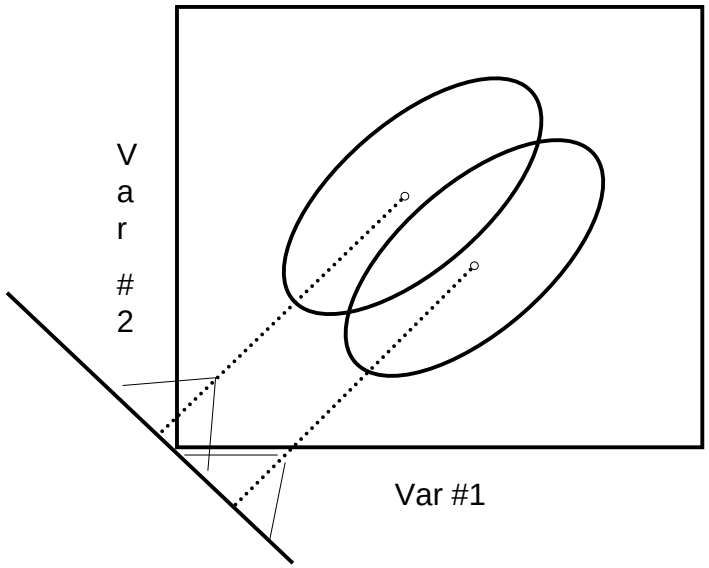
Plot each participant's position in this "2-space", keeping track of group membership. Mark each groups "centroid"

Look at the group difference on each variable, separately.



The dash/dot lines show the mean difference on each variable -- which are small relative to within-group differences, so small Fs

The ldf or MANOVA/Canonidal variate "positioned" to maximize F



In this way, two non-discriminating variables can combine to "work"

Interpreting & Reporting the results of a 2-group ldf Analysis

1. Does the model work ?
   - basic summary statistic is λ (Wilk's lamba) -- smaller is better
   - transformed into X^2 to test H0: of "sphericity"
2. How well does the model work ?
   - λ can be interpreted, with practice`
   - R_c canonical correlation -- like R from multiple regression
     - R_c^2 is between group variance accounted for by ldf
   - "pct of variance" -- tells % of between group variance (100%)
   - % correct reclassification -- results from applying model to assign participants to groups

3. Interpreting the ldf

Three possible bases for interpretation

- unstandardized or raw discriminant weights
  - just like multiple regression weights (but no signif tests)
  - of limited utility because of collinearity
- standardized discriminant weights
  - just like multiple regression β weights
  - useful for "unique contribution" interpretation
- discriminant structure weights
  - correlations between ldf and each variable
    - useful for "descriptive" interpretation

The best (most complete) interpretation will result by combining the information from the standardized and structure weights !!

4. Comparing the bivariate and multivariate group differences

- As with multiple regression, we can have various suppressor effects, such that variables "contribute" to the ldf differently than their bivariate relationship with group membership

5. Determining what the ldf "does for us" -- discrimination

- Consider the group centroids (means) on the ldf - big difference?
  - Centroids will be symmetrical around zero with 2 "=n" grps
- Consider the re-classification results
  - an over-estimate of model's discriminating power (since uses the same participants upon which the model was built)
  - compare model's performance to "baserate" or "chance"
  - look for "asymmetry" -- sometimes one group is easier to identify than the other
  - might employ cross-validation or a hold-out sample to improve the utility of the assessment