

## Multiple-group linear discriminant function

- maximum & contributing ldf dimensions
- concentrated & diffuse ldf structures
- follow-up analyses
- evaluating & reporting k-group ldf

Like ANOVA, ldf can be applied to more than two groups.

- When we have multiple groups there may be an advantage to using multiple discriminant functions to maximally discriminate between the groups.
- That is, we must decide whether the multiple groups “line up” on a single dimension (called a concentrated structure), or whether they are best described by their position in a multidimensional “space” (called a diffuse structure).

Maximum # dimensions for a given analysis:

the smaller of  $\longrightarrow$  # groups - 1  
 $\longrightarrow$  # predictor variables

e.g., 4 groups with 6 predictor variables ? Max # ldfs = \_\_\_\_\_

### “Anticipating” the number of dimensions (ldfs)

By inspecting the “group profiles,” (means of each group on each of the predictor variables) you can often anticipate whether there will be more than one ldf ...

- if the groups have similar patterns of differences (similar profiles) for each predictor variable (for which there are differences), then you would expect a single discriminant function.
- If the groups have different profiles for different predictor variables, then you would expect more than one ldf

Group	Var1	Var2	Var3	Var4
1	10	12	6	8
2	18	12	10	2
3	18	12	10	2

Group	Var1	Var2	Var3	Var4
1	10	12	6	14
2	18	6	6	14
3	18	6	2	7

Concentrated + 0 + -      Diffuse 1st + - 0 0  
 2nd 0 0 - -

Your turn...

Group	Var1	Var2	Var3	Var4
1	21	33	26	68
2	19	34	28	65
3	20	35	12	18

Group	Var1	Var2	Var3	Var4
1	20	33	28	38
2	18	15	26	37
3	42	13	42	38

Group	Var1	Var2	Var3	Var4
1	20	13	28	38
2	18	15	46	36
3	42	43	42	11

Group	Var1	Var2	Var3	Var4
1	23	35	8	38
2	20	36	7	39
3	11	61	2	40

## Determining the number of dimensions (Idfs)

Like other “determinations”, there is a significance test involved

- Each ldf is tested as to whether it “contributes to the model” using the  $X^2$ -test of the  $\lambda$ -value.
- The first ldf will always account for the most between-group variation (have the largest  $X^2$  and  $R_c$ ) -- subsequent ldfs are “orthogonal” (providing independent information), and will account for successively less between group variation.
- If there is a single ldf, then the model is said to have a concentrated structure
- if there are 2 or more ldfs then the model has a diffuse structure
- the distinction between a concentrated and a diffuse structure is considered the “fundamental multivariate question” in a multiple group analysis.



## Follow-up analyses

- Within Psychology, ldf developed in areas of research that traditionally used large samples (e.g., measurement theory & clinical diagnostic research).
- With such large samples, “almost everything is significant”.
- So, an emphasis on “substantial effects” developed - based on “cutoffs” and “relative size” rather than significance tests
  - using % variance to determine if additional ldfs “contribute”
  - .3-.4 cutoff for structure weights when interpreting the ldfs
  - using % classification to discuss “what the model does”
- As ldf was “adopted” into research areas with strong traditions of significance testing, more tests were incorporated into ldf, most commonly “follow-ups”

## There are three major types of follow-ups

- Univariate follow-ups -- abandoning the multivariate analysis, simply describe the results of the ANOVA (with pairwise comparisons) for each of the predictors (DVs)
- Ldf follow-ups -- use the ldf(s) as DVs in ANOVA (with pairwise comparisons) to explicate what which ldfs discriminate between what groups
  - this nicely augments the spatial & re-classification depictions
  - if you have a concentrated structure, it tells you exactly what groups can be significantly discriminated
  - if you have a diffuse structure, it tells you whether the second ldf provides discriminatory power the 1st doesn't

- pairwise ldf follow-ups -- separate ldf analyses for each pair of groups to explicate which variables maximally discriminate between what groups
  - this is just what it sounds like
    - compare groups 1 & 2 then 1 & 3 then 2 & 3
    - interpret the ldf and tell the discriminatory power for each
  - might produce pairwise discriminations not provided by the overall analysis

Different texts/researchers seem to have strong opinions about which of these is the “true multivariate follow-up” or which is “more multivariate”. I'd suggest that usually one of them is a more direct test of the way you have conceptualized your research question or analysis. Trying both probably won't hurt!



## Reporting the Results of a k-group Ldf Analysis

1. Does the model work -- does each possible ldf contribute?
  - $\lambda$  for each ldf transformed into either  $X^2$  to test whether or not that ldf contributes to the model
2. How well does the model work -- ?
  - Overall fit of the model to the data
    - $\lambda$  of the first ldf -- can be interpreted, with practice
    - % correct reclassification -- results from applying full model to assign participants to groups
  - Contribution of each ldf
    - $R_c$  canonical correlation and  $R_c^2$
    - “pct of variance” -- tells % of between group variance attributable to each ldf -- gives good “relative” index

### 3. Interpreting EACH ldf

- The best (most complete) interpretation will result by combining the information from the standardized weights and structure weights !!
- A given variable might contribute to more than one ldf !
  - Consider the structure weights
    - if a variable has a structure weight of .6 for the 1st ldf, that means that the ldf “uses” about 1/3 ( $.6^2=.36$ ) of that variable
    - so 2/3 of that variable is “left over” - possibly to be used to further discriminate between the groups by contributing to another ldf
- Be sure to consider the SIGN of the structure weights when interpreting each.

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 model “does for us” -- discrimination
  - Considering the contribution of each individual ldf
    - Consider the position of the group centroids (means) on that ldf -- which groups are discriminated by which ldf(s)?
  - Consider the functioning of the full model (if diffuse structure)
    - consider the position of the group centroids in the ldf space
    - look at the % reclassification
      - look for “asymmetry” -- sometimes the model will discriminate between some pairs of groups but not between other pairs of groups
      - remember that this % correct is likely to be an overestimate of how well the model will do with new “cases” -- consider a cross-validation or holdout analysis