**A k-group Linear Discriminant Analysis & Follow-ups**

Aside from the occasional suppressor effect or asymmetrical classification, a 2-group ldf analysis rarely tells us more than does the set of F-tests. So, we will spend our time on the more interesting and important k-group models.

* The grouping variable is **jobcat**
* Discriminating variables include **educ, salary, salbegin, jobtime, prevexp, minority**
1. Means (standard deviations) and ANOVAs for the discriminating variables.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | educ | salary | salbegin | jobtime | prevexp | minority |
| Group | Mean (std) | Mean (std) | Mean (std) | Mean (std) | Mean (std) | Mean (std) |
| Clerical |  |  |  |  |  |  |
| Custodial |  |  |  |  |  |  |
| Manager |  |  |  |  |  |  |
| F (df, dferror) |  |  |  |  |  |  |
| p |  |  |  |  |  |  |

1. Consider the table of means you constructed above. Inspecting the patterns of mean differences across the groups. Do you expect a concentrated or a diffuse discriminant structure for these variables? Explain your answer.

Obtain the discriminant model (using all the predictors).

1. From the discriminant analysis 1st 🡺 X² (\_\_) = \_\_\_\_\_\_, p = \_\_\_\_\_\_ R² -cannonical = \_\_\_\_\_\_

 2nd 🡺 X² (\_\_) = \_\_\_\_\_\_, p = \_\_\_\_\_\_ R² -cannonical = \_\_\_\_\_\_

1. Structure weights for each variable on each ldf. Underline the contributors to each ldf (use +/- .30 as the cutoff).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | educ | salary | salbegin | jobtime | prevexp | minority |
| ldf #1 |  |  |  |  |  |  |
| ldf #2 |  |  |  |  |  |  |

1. Classification table -- include both % and (number) for each cell

|  |  |
| --- | --- |
| **Original group membership** | Predicted group membership |
| Clerical | Custodial | Manager |
| Clerical |  |  | % correct classification = \_\_\_\_\_\_\_ |
| Custodial |  |  |  |
| Manager |  |  |  |

1. Centroids for each group on each ldf

|  |  |  |
| --- | --- | --- |
|  | ldf #1 | ldf #2 |
| Clerical |  |  |
| Custodial |  |  |
| Manager |  |  |

1. Make the graphical representation of these data. Position each group & label the ldfs with contributing discriminators.
2. Considering the pattern of the centroids, as shown in the graphical presentation, what between-group discrimination is contributed by each ldf?"

ldf #1 –

ldf #2 –

1. Write it up, following the example in the handout.

k-group analyses often reveal interesting patterns of similarities and differences among the groups, but the ldf output leaves several statistical questions about group differences unanswered. The three different kinds of follow-ups all address different versions of the question, “How are the groups different?”

**Bivariate Follow-ups:**

1. For each predictor, use “Post Hoc” comparisons to get the pairwise comparisons using LSD. Use <, > or = to represent the result of the LSD pairwise comparisons for each predictor. Use = for all pairwise comparison of predictors with non-significant F-tests.

|  |  |  |  |
| --- | --- | --- | --- |
| **Predictor** | **Cleric vs Custodial** | **Cleric vs Manager** | **Custodial vs, Manager** |
| educ  |  |  |  |
| salary  |  |  |  |
| salbegin  |  |  |  |
| jobtime  |  |  |  |
| prevexp  |  |  |  |
| minority  |  |  |  |

1. Describe which groups are differentiated by which variables.
2. Describe which variables differentiate which groups.
3. Which description do you like best? Why?

**LDF follow-ups:**

1. Rerun the ldf analysis asking SPSS to save the linear discriminant scores. Then use discriminant scores from each significant ldf as dependent variables in an ANOVA. Be sure to ask for LSD Post Hoc analyses.

|  |  |  |
| --- | --- | --- |
|  | **ldf #1** | **ldf #2** |
| Group | Mean (std) | Mean (std) |
| Clerical |  |  |
| Custodial |  |  |
| Manager |  |  |
| F (df, dferror) |  |  |
| p |  |  |

1. Use <, > or = to represent the result of the LSD pairwise comparisons for each ldf. Use = for all pairwise comparison of ldfs with non-significant F-tests.

|  |  |  |  |
| --- | --- | --- | --- |
| **Predictor** | **Cleric vs Custodial** | **Cleric vs Manager** | **Custodial vs, Manager** |
| ldf #1 |  |  |  |
| ldf #2 |  |  |  |

1. Describe which groups are differentiated by which variables.
2. Describe which variables differentiate which groups.
3. Which description do you like best? Why?

**Pairwise LDF Analyses**

Run an ldf using just the Custodial and Manager groups.

Wilks' Lambda = \_\_\_\_\_\_ X² = \_\_\_\_\_\_ df = \_\_\_ p = \_\_\_\_\_\_ R² (not R) = \_\_\_\_\_\_

% correct reclassification for 1st group (custodial) = \_\_\_\_\_\_ 2nd group (manager) = \_\_\_\_\_\_

|  |  |  |
| --- | --- | --- |
| **Predictor** | **ldf #1****structure weight** | **ldf #1****β weight** |
| educ  |  |  |
| salary  |  |  |
| salbegin  |  |  |
| jobtime  |  |  |
| prevexp  |  |  |
| minority  |  |  |

1. Which variables separate these two groups? (Use structure weights)
2. Are either of these groups reclassified better than they were in the k-group model?
3. Are the variables that differentiate these groups “surprising” when you look back at the graphic of the full model? That is, do you have a different understanding of “how these groups differ” based upon this analysis than you had after interpreting the full model?