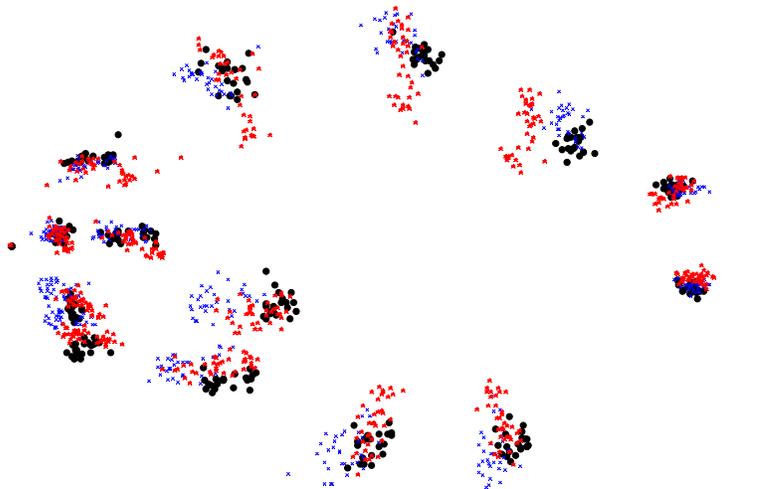


IMP: CVAGen6

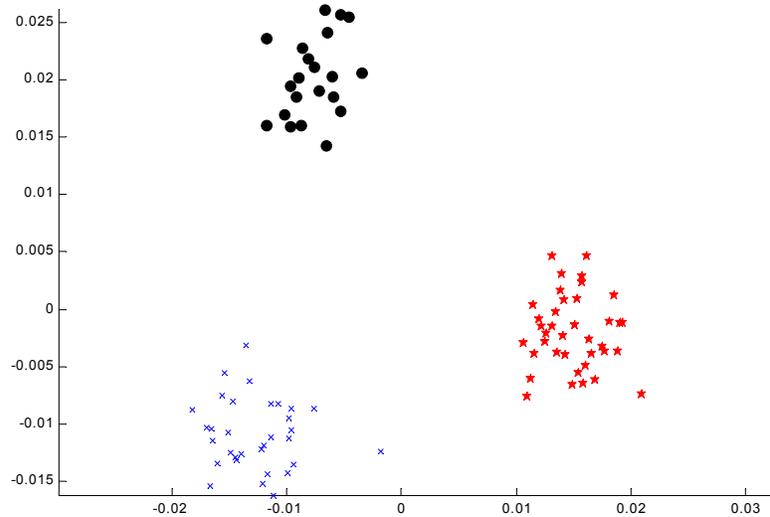
*H. David Sheets, Dept. of Physics, Canisius College, Buffalo, NY 14208,
sheets@canisius.edu*

Introduction:

This is a Canonical Variates Analysis program for the analysis of shape, based on partial warp scores, and is part of the IMP software series. Canonical Variates Analysis is a method of finding the set of axes that allows for the greatest possible ability to discriminate between two or more groups. The program computes partial warp scores to a common reference and then does a Manova followed by a CVA. It determines how many distinct CVA axes there are in the data at a $p=0.05$ level of significance, and computes the canonical variates scores of all the specimens entered. It also uses Mahalanobis distances to assign all the loaded specimens to one of the groups loaded. The image below shows the Bookstein coordinates of variables from the example file, threepir.txt, which has three groups of piranha.



There are two distinct canonical variates axes for this set of three groups, and the canonical variate scores of these same specimens plotted along the canonical variates axes shown below illustrate the clear ability to separate these three groups.



Using CVAGen6

I haven't had time to write a complete manual for CVAGen to date, just what is presented here. The operation of the program is very similar to that used by PCAGen, so obtain that program and it's manual if you have not done so already.

File Formats

CVAGen6 uses the same file formats as the PCAGen6 program, and essentially the same control format. If you haven't used PCAGen6, obtain it, read the manual and run it before proceeding with CVAGen6. The grouplist is required for CVAGen6 (unlike the optional group list in PCAGen6).

Sample Files

The files threepir.txt and threepirgrp.txt are a text file of three species of piranha and the

associated group file necessary to run CVAGen6. Try these files to get a sense of how the program operates.

Tests of significance used in CVAGen

The tests of significance of the canonical variate axes in CVAGen are all based on the Wilk's lambda (λ) value, which is the sum of squares within groups divided by the total sum of squares within and between groups

$$\lambda = |W| / |W+B| = |W| / |T|$$

Bartlett's test then uses the following test statistic

$$X^2 = -(w - (p-b+1)/2) \ln \lambda$$

where X^2 has an approximately chi-squared distribution, w is the degrees of freedom for the within group sum of squares, b is the degrees of freedom for the within group sum of squares and p is the number of variables, to determine if there are $g = b+1$ distinct groups. The degrees of freedom within is $w=n-b$, where n is the total number of samples and g is the number of groups.

The results of these tests will appear onscreen in a new window (the Auxillary Results Box) when you load the group list file. The contents of this window may be appended to a text file, by using the Append Results to File button on the bottom of the window. Note

that if the filename you specify does not already exist, the Append Results to File Button will create a new text file to put the results in. You might sometimes get a warning message about overwriting a file. Don't worry about this message when using this option, it won't really overwrite the file, but rather will append it.

Assignment of specimens to groups

A simple Mahalanobis distance-based approach is then used to determine which group each specimen belongs to, based on the canonical variate scores. The predicted group membership of each specimen based on the CVA scores is determined by assigning each specimen to the group whose mean is closest (under the Mahalanobis distance) to the specimen. Note that this is not a sophisticated approach to partitioning the space determined by the CVA axis, but it was algorithmically simple. I would like to find something better than this, but for now, it is better than nothing.

CVAGen produces a matrix showing group assignments, the original (user-assigned) groups are rows of the matrix, the new groups are along the columns of the matrix. The rows and columns are labeled with the group codes from the group list file input into the program. Study of this matrix allows the user to assess how effective the CVA is at separating groups, and which groups were not successfully separated.

Output from Sample files

Below are the results obtained from the sample files. The first section lists the Wilk's Lambda value, chi-squared value and degrees of freedom as well as the p score for the various number of CVA axes (not axis, gotta fix that!).

Results from CVA/Manova

Axis 1 Lambda= 0.0031 chisq=430.9308 df=56 p<2.22045e-016

Axis 2 Lambda= 0.0684 chisq=199.7954 df=27 p<2.22045e-016

Groupings from CVA-Distance Based

Original Groups along rows, CVA groups along columns

-

- 0 1 2 3

- 1 21 0 0

- 2 0 32 0

- 3 0 0 38

The second section shows the assignments of specimens to various groups, based on the Mahalanobis distance in the space defined by the significant CVA axes. The original groups are along the rows of the matrix, the predicted group (from the CVA) is along the columns. In this case, all specimens were assigned to the correct groups.

In the second example shown below, there were 9 groups and 8 significant axes.

Membership assignment was always correct for groups 1 and 3, but specimens in Group two were also assigned to groups 7 (4 specimens) and 8 (1 specimen). One specimen from group 4 was assigned to group 5, one specimen from group 5 was incorrectly assigned to group 6, and so forth.

Groupings from CVA-Distance Based

Original Groups along rows, CVA groups along columns

-										
-	0	1	2	3	4	5	6	7	8	9
-	1	31	0	0	0	0	0	0	0	0
-	2	0	16	0	0	0	0	4	1	0
-	3	0	0	32	0	0	0	0	0	0
-	4	0	0	0	45	1	0	0	0	0
-	5	0	0	0	0	37	1	0	0	0
-	6	0	0	0	0	7	33	0	0	0
-	7	0	9	0	0	0	0	77	3	0
-	8	0	1	0	0	0	0	1	32	0
-	9	0	0	0	0	0	0	0	0	37

Displaying Results

CVA Axis

The program can plot out the deformation implied by the CVA axes or the change in shape obtained by regressing the shape on the CVA axis scores. There is a pushbutton labeled *Regr?* on the screen in the Display CVA Axes section that control which you will see, if this button is pushed (black) you will see the regression deformation. The regression shows you all the changes in shape that are correlated with the CVA axis score, the CVA axis itself does not show all the correlated change in shape.

Show Deformation Implied by CVA

These options let you place two markers on the scatter plot of CVA scores, and then shows you the deformation implied by the markers, relative to the reference form. This allows you to see the effects of differences along two CVA axes at the same time.