Correspondence Analysis

Correspondence analysis is a method for examining the relationship between two sets of categorical variables much as in a Chi-Squared analysis of a two-way contingency table. In fact, a typical chi-squared analysis is completed as part of this procedure. In addition, visualization of the relationships among the columns or rows of the analysis is performed in a manner similar to factor analysis. The data analyzed in the visualization is the table of relative proportions, that is, the original frequency values divided by the sum of all frequencies. The relative proportions of the row sums and the column sums are termed the "masses" of the rows or columns.

The method used to analyze the relative proportions involves what is now called the "Generalized Singular Value Decomposition" or more simply the generalized SVD. This method obtains roots and vectors of a rectangular matrix by decomposing that matrix into three portions: a matrix of left singular column vectors (A) that has n rows and q columns ($n \geq q$), a square diagonal matrix with q rows and columns of singular values (D), and a transposed matrix (B') that is m x q in size of right generalized singular vectors ($m = q-1$). Completing this analysis involves several steps. The first is to obtain the (regular) SVD analysis of a matrix Q defined as $D_r^{-1/2}\ P\ D_c^{-1/2}$ where D_r and D_c are diagonal matrices of row and column relative proportions and P is the matrix of relative proportions. The SVD of Q gives
$Q = U\ D\ V'$ where D is the desired diagonal matrix of eigenvalues and $U'U = V'V = I$. It should be noted that the first of the q roots is trivial and to be ignored. At this point we obtain $A = D_r^{1/2}$ Uand $B = D_c^{1/2}\ V$. The results of this SVD analysis is available on the output. Now $P = ADB'$. The row coordinates F and column coordinates G are then computed according to the table below:

Analysis Choice	Button Selected	Row Coordinates	Column Coordinates
Row Profile	Row	$F = D_r^{-1}AD$	$G = D_c^{-1}B$
Column Profile	Column	$F = D_r^{-1}A$	$G = D_c^{-1}BD$
Both Profiles	Both	$F = D_r^{-1}AD$	$G = D_c^{-1}BD$

If Row profiles are computed, the row coordinates are weighted centroids of the column coordinates and the inertias D^2 refer only to the row points. If the column profiles are computed, the column coordinates are weighted eentroids of the row coordinates and the inertias D^2 refer only to the column points. If both profiles are selected, neither row or column coordinates are weighted centroids of the other but the inertias D^2 refer to both sets of points. The q-1 inertias are plotted in a manner similar to a scree plot of roots in a factor analysis. The total inertia is, in fact, the chi-squared statistic divided by the total of all cell frequencies.

You may elect to plot the coordinates for any two pairs of coordinates. This will provide a graphical representation of the separation of the row or column categories similar to a plot of variables in a discriminant function analysis or factors in a factor analysis. A way of looking at correspondence analysis is to consider it as a method for decomposing the overall inertia by identifying a small number of dimensions in which the deviations from the expected values can be represented. This is similar to factor analysis where the total variance is decomposed so as to arrive at a lower dimensional representation of variables.