Reveal Underlying Relationships in Categorical Data

Unleash the full potential of your data through perceptual mapping, optimal scaling, and dimension reduction techniques. With SPSS Categories' sophisticated procedures in your analytical toolbox, you aren't hampered by categorical or high-dimensional data. SPSS Categories provides you with everything you need to analyze and interpret multivariate data and their relationships more completely.

You can rely on SPSS Categories whenever you need to:

- Visually interpret how rows and columns relate in large tables of counts, ratings, or rankings
- Work with and understand ordinal and nominal data with procedures similar to conventional regression, principal components, and canonical correlation
- Perform regression using ordered or unordered categorical outcome variables
- Determine how closely customers perceive your products when compared to other products in your or your competitors' offerings
- Understand what characteristics consumers relate most closely to your product or brand

Turn your qualitative variables into quantitative ones

Use SPSS Categories' advanced optimal scaling procedures to assign units of measurement and zero-points to your categorical data. This opens up a new set of statistical functions by allowing you to perform analyses on variables of mixed measurement levels (nominal, ordinal, and numerical variables).

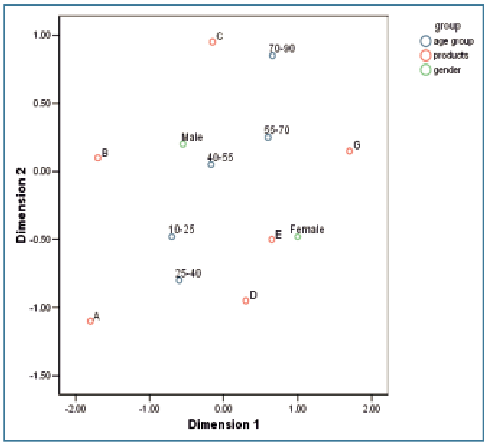
Correspondence and multiple correspondence analyses enable you to numerically evaluate the similarities between two or more nominal variables.

You can better understand your data using categorical principal components analysis. Summarize your data using important components based on variables of mixed measurement levels. You also can incorporate variables of different measurement levels into sets and analyze them using nonlinear canonical correlation analysis.

Discover underlying relationships—graphically

SPSS Categories' dimension reduction techniques enable you to go beyond unwieldy tables to clearly see relationships in your data using revealing perceptual maps and biplots. Perceptual maps are high-resolution summary charts that graphically display similar variables or categories close to each other. They provide you with unique insight into relationships between more than two categorical variables. Biplots enable you to look at the relationships among cases, variables, and categories. For example, examine customers, brands, and characteristics.

Whatever types of categories you study—market segments, subcultures, political parties, or biological species—optimal scaling procedures free you from the restrictions associated with two-way tables, placing the relationships among your variables in a larger frame of reference. You can see a map of your data—not just a statistical report.



The data shown at left are a 2x5x6 table containing information on two genders, five age groups, and six products. This biplot shows the results of a two-dimensional multiple correspondence analysis of the table. Notice that products such as "A" and "B" are chosen at younger ages and by males, while products such as "G" and "C" are preferred at older ages.

How you can use SPSS Categories

Categorical regression (CATREG) predicts the values of an ordered or unordered categorical outcome variable from a combination of categorical predictor variables.

Regression with optimal scaling quantifies categorical predictor variables so that they have maximally strong relationships with nominal, ordinal, or numerical outcome variables.

You can use regression with optimal scaling to describe, for example, how job satisfaction depends on job category, geographic region, and the amount of work-related travel. You may find that high levels of satisfaction apply to managers and to employees who have little work-related travel.

You could then use the regression equation that results to predict job satisfaction for any combination of the categories using the three predictor variables.

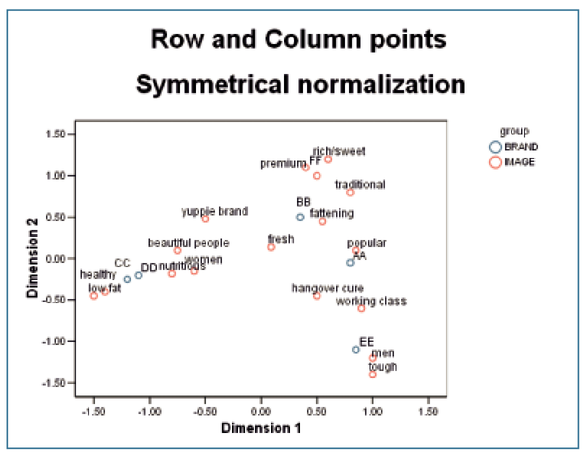
Correspondence analysis (CORRESPONDENCE) enables you to analyze two-way tables whose cells contain some measurement of correspondence between the rows and columns. A very common type of correspondence table is a crosstabulation, where the cells contain frequency counts.

Correspondence analysis describes the relationship between two nominal variables in a perceptual map, while simultaneously describing the relationship between the categories of the variables.

Better Understand Consumer Perceptions

Market researchers in South Australia wanted to better understand how consumers perceived the six iced coffee brands in the market. They asked consumers to rate each of the brands (denoted AA to FF) on 16 different categorical attributes. The 96-cell table that resulted was too large to clearly see the relationships between the brands and the perceived attributes.

The market researchers used the correspondence procedure in SPSS to identify the two strongest underlying factors in the relationships between the brands and attributes. By assigning each brand and attributing a specific number within each dimension, the information was displayed in an easily understood chart, commonly called a perceptual map. For example, it is clear from the chart that Brand AA is the brand most closely identified by the market with the "popular" attribute, which makes sense since it is the category leader. Similarly, researchers can quickly identify that consumers who are interested in healthy and low-fat products perceive CC and DD with greater regard, while FF is perceived as a rich, sweet premium brand.*



Researchers studied the consumer perceptions of six iced coffee brands sold in South Australia. Brands are denoted AA to FF and are characterized by various categorical attributes, such as "healthy." The correspondence procedure in SPSS produced the correspondence map shown here.

* *Source for data and example: Kennedy, R., C. Riquier, and Byron Sharp. 1996. "Practical Applications of Correspondence Analysis to Categorical Data in Market Research,"* Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 5, No. 1, pp. 56-70.*

For example, you can use correspondence analysis to graphically display the relationship between marital status and level of general happiness. You might find that widowed and separated people have very similar levels of happiness, while married people have much higher levels of happiness.

Multiple correspondence analysis (MULTIPLE CORRESPONDENCE) is used to analyze multivariate categorical data where all the variables are analyzed at the nominal level (unordered categories). This analysis is similar to correspondence analysis; however, it allows you to use more than two variables.

Multiple correspondence analysis simultaneously describes the relationships between cases and categories by displaying them in a low-dimensional map.

For example, multiple correspondence analysis can graphically display the relationship among favorite television show, age group, and gender. From this, you can find the age and gender demographics that gravitate to each show while simultaneously revealing which shows are most similar.

Categorical principle components analysis (CATPCA) uses optimal scaling to generalize principal components analysis to accommodate variables of mixed measurement levels. Unlike when using multiple correspondence analysis, you are able to specify an analysis level (nominal, spline nominal, ordinal, spline ordinal, or numeric) on a variable-by-variable basis.

CATPCA displays the variables in terms of their categories. If the variable is ordinal or numeric, the category points are on a straight line. The direction of the line displays the correlation of the variable with the principal components. If the variable is nominal, you can also choose to display the categories as midpoints of cases (as in multiple correspondence analysis).

For example, you can display the relationships between different brands of cars (the cases) and a number of their characteristics (price, weight, mileage, etc.). Describe cars by their class (compact, midsize, convertible, SUV, etc.), and CATPCA uses these classifications to group the points for the cars. If consumers have ranked a number of midsize cars, you can analyze rankings with ordinal treatment of the variables (the consumers), and with cars as cases in the analysis.

Nonlinear canonical correlation analysis (OVERALS) uses optimal scaling to generalize canonical correlation analysis to accommodate variables of mixed measurement levels.

This type of analysis enables you to compare multiple sets of variables to one another in the same graph, after removing the correlation within sets.

For example, sets could be characteristics of products (such as soups) in a taste study. The judges represent the variables within the sets; the soups are the cases. OVERALS then averages the judges after removing the correlations, and combines the different characteristics to display the relationships between the soups. Alternatively, each judge may have used his or her own characteristics (the variables) to judge the soups, and each judge forms a set. In the latter case, OVERALS averages over the characteristics after removing the correlations, and then combines the scores for the different judges.

Multidimensional scaling (PROXSCAL) performs multidimensional scaling of one or more matrices with similarities or dissimilarities (proximities). Alternatively, you can compute distances between cases in multivariate data as input to PROXSCAL.

PROXSCAL displays proximities as distances in a map in order for you to gain spatial understanding of how objects relate. In the case of multiple proximity matrices, PROXSCAL analyzes the communalities and the differences between them.

For example, you can use PROXSCAL to display the similarities between different cola flavors according to consumers in various age groups. You might find that young people emphasize differences between traditional and new flavors, while adults emphasize diet versus non-diet colas.

Features*

Statistics

CATREG

- Categorical regression analysis via optimal scaling
  - Specify the optimal scaling level at which you want to analyze each variable. Choose from: Spline ordinal (monotonic), spline nominal (nonmonotonic), ordinal, nominal, multiple nominal, or numerical.
  - Discretize continuous variables or convert string variables to numeric integer values by multiplying, ranking, or grouping values into a preselected number of categories according to an optional distribution (normal or uniform), or by grouping values in a preselected interval into categories. The ranking and grouping options can also be used to recode categorical data.
  - Specify how you want to handle missing data. Impute missing data with the variable mode or with an extra category, or use listwise exclusion.
  - Specify objects to be treated as supplementary
  - Specify the method used to compute the initial solution
  - Control the number of iterations
  - Specify the convergence criterion
  - Plot results
    - Transformation plots (optimal category quantifications against category indicators)
    - Residual plots
  - Add transformed variables, predicted values, and residuals to the working data file
  - Print results
    - Multiple R, R^2, and adjusted R^2
    - Standardized regression coefficients, standard errors, zero-order correlation, part correlation, partial correlation, Pratt's relative importance measure for the transformed predictors, tolerance before and after transformation, and F statistics
    - Table of descriptive statistics including marginal frequencies, transformation type, number of missing values, and mode
    - Iteration history
    - Tables for fit and model parameters: ANOVA table with degrees of freedom according to optimal scaling level, model summary table with adjusted RSQ for optimal scaling, t-values, significance levels, a separate table with the zero-order, part and partial correlation, and the importance and tolerance before and after transformation
    - Correlations of the transformed predictors and Eigenvalues of the correlation matrix
    - Correlations of the original predictors and Eigenvalues of the correlation matrix
    - Category quantifications
  - Write discretized and transformed data to an external data file

CORRESPONDENCE

- Correspondence analysis
  - Input data as a case file or directly as table input
  - Specify the number of dimensions of the solution
  - Choose from two distance measures: Chi-square distances for correspondence analysis or Euclidean distances for biplot analysis types
  - Choose from five types of standardization: Remove row means, remove column means, remove row-and-column means, equalize row totals, or equalize column totals
  - Five types of normalization: Symmetrical, principal, row principal, column principal, and customized
  - Print results
    - Correspondence table
    - Summary table: Singular values, inertia, proportion of inertia accounted for by the dimensions, cumulative proportion of inertia accounted for by the dimensions, confidence statistics for the maximum number of dimensions, row profiles, and column profiles
    - Overview of row and column points: Mass, scores, inertia, contribution of the points to the inertia of the dimensions, and contribution of the dimensions to the inertia of the points
    - Row and column confidence statistics: Standard deviations and correlations for active row and column points
    - Permuted table: Table with rows and columns ordered by row and column scores for a given dimension
  - Plot results: Row scores, column scores, and biplot (joint plot of a row or column score)
  - Write row scores, column scores, and confidence statistics (variances and covariances) to an external data file

MULTIPLE CORRESPONDENCE

- □ Multiple correspondence analysis (replaces HOMALS, which was included in versions prior to SPSS Categories 13.0)
  - Specify variable weights
  - Discretize continuous variables or convert string variables to numeric integer values by multiplying, ranking, or grouping values into a preselected number of categories according to an optional distribution (normal or uniform), or by grouping values in a preselected interval into categories. The ranking and grouping options can also be used to recode categorical data.
  - Specify how you want to handle missing data. Exclude only the cells of the data matrix without valid value, impute missing data with the variable mode or with an extra category, or use listwise exclusion.
  - Specify objects and variables to be treated as supplementary (full output is included for categories that occur only for supplementary objects)

** Features subject to change based on final product release.* □ *Symbol indicates a new feature.*

- Specify the number of dimensions in the solution
- Specify a file containing the coordinates of a configuration and fit variables in this fixed configuration
- Choose from five normalization options: Variable principal (optimizes association between variables), object principal (optimizes distances between objects), symmetrical (optimizes relationship between objects and variables), independent, or customized (user-specified value allowing anything in between variable principal and object principal normalization)
- Control the number of iterations
- Specify convergence criterion
- Print results
  - Model summary
  - Iteration statistics and history
  - Descriptive statistics (frequencies, missing values, and mode)
  - Discrimination measures per variable and per dimension
  - Category quantifications (centroid coordinates), mass, inertia of the categories, contribution of the categories to the inertia of the dimensions, and contribution of the dimensions to the inertia of the categories
  - Correlations of the transformed variables and the Eigenvalues of the correlation matrix for each dimension
  - Correlations of the original variables and the Eigenvalues of the correlation matrix
  - Object scores
  - Object contributions: Mass, inertia, contribution of the objects to the inertia of the dimensions, and contribution of the dimensions to the inertia of the objects
- Plot results
  - Category plots: Category points, transformation (optimal category quantifications against category indicators), residuals for selected variables, and joint plot of category points for a selection of variables
  - Plot of the object scores
  - Plot of discrimination measures
  - Biplot of objects and centroids of selected variables
- Add transformed variables and object scores to the working data file
- Write discretized data, transformed data, and object scores to an external data file

CATPCA

- Categorical principal components analysis via optimal scaling
  - Specify the optimal scaling level at which you want to analyze each variable. Choose from: Spline ordinal (monotonic), spline nominal (nonmonotonic), ordinal, nominal, multiple nominal, or numerical.
  - Specify variable weights
  - Discretize continuous variables or convert string variables to numeric integer values by multiplying, ranking, or grouping values into a preselected number of categories according to an optional distribution (normal or uniform), or by grouping values in a preselected interval into categories. The ranking and grouping options can also be used to recode categorical data.
  - Specify how you want to handle missing data. Exclude only the cells of the data matrix without valid value, impute missing data with the variable mode or with an extra category, or use listwise exclusion.
  - Specify objects and variables to be treated as supplementary (full output is included for categories that occur only for supplementary objects)
  - Specify the number of dimensions in the solution
  - Specify a file containing the coordinates of a configuration and fit variables in this fixed configuration
  - Choose from five normalization options: Variable principal (optimizes association between variables), object principal (optimizes distances between objects), symmetrical (optimizes relationship between objects and variables), independent, or customized (user-specified value allowing anything in between variable principal and object principal normalization)
  - Control the number of iterations
  - Specify convergence criterion
  - Print results
    - Model summary
    - Iteration statistics and history
    - Descriptive statistics (frequencies, missing values, and mode)
    - Variance accounted for per variable and per dimension
    - Component loadings
    - Category quantifications and category coordinates (vector and/or centroid coordinates) for each dimension
    - Correlations of the transformed variables and the Eigenvalues of the correlation matrix
    - Correlations of the original variables and the Eigenvalues of the correlation matrix
    - Object (component) scores
  - Plot results
    - Category plots: Category points, transformation (optimal category quantifications against category indicators), residuals for selected variables, and joint plot of category points for a selection of variables
    - Plot of the object (component) scores
    - Plot of component loadings

- Biplots: Of objects and loadings, objects and centroids, and loadings and centroids
  - Triplot of objects, loadings, and centroids
  - Plot of the centroids of a variable projected on the vector of each of a selection of variables
- Select the variables to include in biplots and triplot
- Add transformed variables, object scores, and the approximation of the transformed variables to the working data file
- Write discretized data, transformed data, object scores, and the approximation of the transformed variables to an external data file

OVERALS

- Generalized canonical correlation analysis via optimal scaling
  - Specify the optimal scaling level at which to analyze each variable. Choose from: Ordinal, nominal, multiple nominal, or numerical.
  - Specify the method used to compute the initial configuration
  - Specify the number of sets, and which variables belong in each set
  - Specify the number of dimensions
  - Specify the maximum number of iterations
  - Specify the convergence criterion
  - Print results
    - Marginal frequencies for variables in the analysis
    - Iteration history
    - Multiple fit
    - Single fit
    - Single loss per variable
    - Centroids
    - Projected centroids
    - Object scores
    - Category quantifications
    - Single and multiple coordinates
    - Weights
    - Component loadings
  - Specify category plots: Joint plot of category points, joint plot of centroids, and transformations (optimal category quantifications against category indicators)
    - Plot of component loadings
    - Plot of object scores
  - Write category quantifications, coordinates, centroids, weights, and component loadings to an external data file

PROXSCAL

- Multidimensional scaling analysis
  - Read one or more square matrices of proximities, either symmetrical or asymmetrical
  - Read weights, initial configurations, fixed coordinates, and independent variables
  - Treat proximities as ordinal (nonmetric) or numeric (metric); ordinal transformations can treat tied observations as discrete or continuous
  - Specify multidimensional scaling with three individual differences models, as well as the identity model
  - Specify fixed coordinates or independent variables to restrict the configuration. Additionally, specify the transformations (numerical, nominal, ordinal, and splines) for independent variables.
  - Specify output that includes the original and transformed proximities, history of iterations, individual space weights, fitted distances, and decomposition of stress
  - Plot results: Stress plots, common space scatterplots, individual space weight scatterplots, individual spaces scatter-plots, transformation plots, Shepard residual plots, independent variables transformation plots, and correlation plots

System requirements

- Software: SPSS Base 13.0
- Minimum free drive space: 1MB
- Other system requirements vary according to platform

SPSS

To learn more, please call Technologies4Targeting Ltd +44 (0)1733 890790 or visit www.tech4t.co.uk/spss