Reveal Relationships in Categorical Data

Unleash the full potential of your data through perceptual mapping, optimal scaling, preference scaling, and dimension reduction techniques. SPSS Categories provides you with all the tools you need to obtain clear insight into complex categorical and high-dimensional data.

With SPSS Categories, you can visually interpret data and see how rows and columns relate in large tables of counts, ratings, or rankings. This gives you the ability to:

- Work with and understand ordinal and nominal data using procedures similar to conventional regression, principal components, and canonical correlation
- Perform regression using nominal or ordinal categorical predictor or outcome variables

For example, use SPSS Categories to understand which characteristics consumers relate most closely to your product or brand, or to determine customer perception of your products, compared to other products that you or your competitors offer.

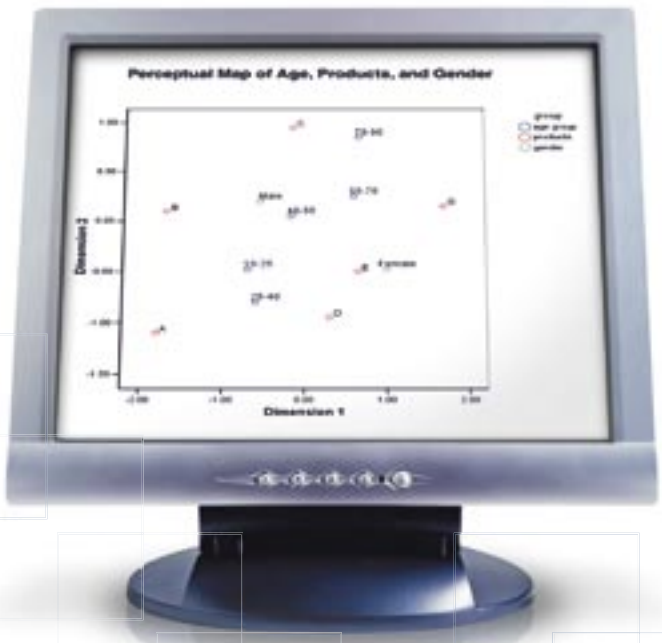
Turn your qualitative variables into quantitative ones

The advanced procedures available in SPSS Categories enable you to perform additional statistical operations on categorical data.

Use SPSS Categories' 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—on nominal, ordinal, and numeric variables, for example.

SPSS Categories' ability to perform correspondence and multiple correspondence analyses helps you numerically evaluate similarities between two or more nominal variables in your data.

And, with its principal components analysis procedure, you can summarize your data according to important components. Or incorporate variables of different measurement levels into sets and then analyze them by using nonlinear canonical correlation analysis.



Graphically display underlying relationships

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.

SPSS Categories' dimension reduction techniques enable you to go beyond unwieldy tables. Instead, you can clarify relationships in your data by using 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, you can define relationships between products, customers, and demographic characteristics.

Now, with the new preference scaling procedure, you can further visualize relationships among objects. The breakthrough algorithm on which this procedure is based enables you to perform non-metric analyses for ordinal data and obtain meaningful results.

How you can use SPSS Categories

The following procedures are available to add meaning to your data analyses.

Categorical regression (CATREG) predicts the values of a nominal, ordinal, or numerical outcome variable from a combination of categorical predictor variables that the procedure quantifies through optimal scaling techniques.

You can use regression with optimal scaling to describe, for example, how job satisfaction relates to job category, geographic region, and the amount of work-related travel.

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

SPSS Categories displays relationships among nominal variables in a perceptual map, a visual presentation that also shows the relationships among the categories of the variables.

Multiple correspondence analysis (MULTIPLE CORRESPONDENCE) is used to analyze multivariate categorical data. It differs from correspondence analysis in that it allows you to use more than two variables in your analysis. With this procedure, all the variables are analyzed at the nominal level (unordered categories).

For example, you can use multiple correspondence analysis to explore relationships between favorite television show, age group, and gender. By examining a low-dimensional map created with SPSS Categories, you could see which groups gravitate to each show while also learning which shows are most similar.

Categorical principal components analysis (CATPCA) uses optimal scaling to generalize the principal components analysis procedure so that it can accommodate variables of mixed measurement levels. It is similar to multiple correspondence analysis, except that you are able to specify an analysis level on a variable-by-variable basis.

For example, you can display the relationships between different brands of cars and characteristics such as price, weight, fuel efficiency, etc. Alternatively, you can describe cars by their class (compact, midsize, convertible, SUV, etc.), and CATPCA uses these classifications to group the points for the cars. SPSS Categories displays results in a low-dimensional map that makes it easy to understand relationships.

Nonlinear canonical correlation analysis (OVERALS) uses optimal scaling to generalize the canonical correlation analysis procedure so that it can 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, you might analyze characteristics of products, such as soups, in a taste study. The judges represent the variables within the sets while the soups are the cases. OVERALS averages the judges' evaluations, after removing the correlations, and combines the different characteristics to display the relationships between the soups. Alternatively, each judge may have used a separate set of criteria to judge the soups. In this instance, each judge forms a set and OVERALS averages the criteria, after removing the correlations, and then combines the scores for the different judges.

Multidimensional scaling (PROXSCAL) performs multidimensional scaling of one or more matrices containing 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 a spatial understanding of how objects relate. In the case of multiple proximity matrices, PROXSCAL analyzes the commonalities and plots the differences between them.

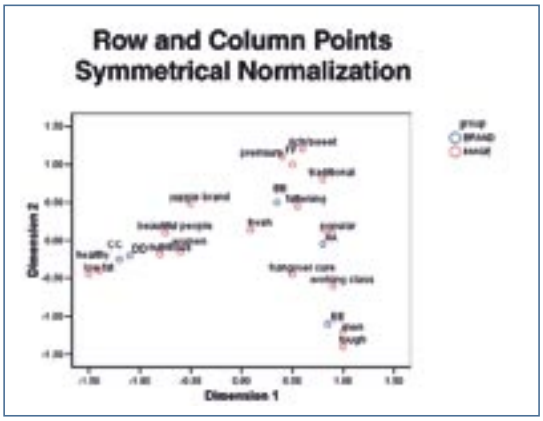
For example, you can use PROXSCAL to display the similarities between different cola flavors preferred by consumers in various age groups. You might find that young people emphasize differences between traditional and new flavors, while adults emphasize diet versus non-diet colas.

Preference scaling (PREFSCAL) visually examines relationships between variables. Preference scaling performs multidimensional unfolding on two sets of objects in order to find a common quantitative scale.

This enables you to find clusters among variables. For example, if a group of drivers rated 26 models of cars on ten attributes on a six-point scale, you could find clusters showing which models were similar, and which attributes were associated with them.

Better Understand Consumer Perceptions

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



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.

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. 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.*

* *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.*

Features

Statistics

CATREG

- Categorical regression analysis through 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, either as:
    - Transformation plots (optimal category quantifications against category indicators)
    - Residual plots
  - Add transformed variables, predicted values, and residuals to the working data file
  - Print results, including:
    - Multiple R, R^2, and adjusted R^2 charts
    - 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 R^2 for optimal scaling, *t* values and 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, including:
    - 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)
  - 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 associations between variables), object principal (optimizes distances between objects), symmetrical (optimizes relationships 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, including:
    - Model summary
    - Iteration statistics and history

- Descriptive statistics (frequencies, missing values, and mode)
- Discrimination measures by variable and 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, creating:
- 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
- Object scores
- Discrimination measures
- Biplots 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 through 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 associations between variables), object principal (optimizes distances between objects), symmetrical (optimizes relationships 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, including:
    - Model summary
    - Iteration statistics and history
    - Descriptive statistics (frequencies, missing values, and mode)
    - Variance accounted for by variable and 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, creating:
    - Category plots: Category points, transformations (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

PREFSCAL (syntax only)

- Visually examine relationships between variables in two sets of objects in order to find a common quantitative scale
  – Read one or more rectangular matrices of proximities
  – Read weights, initial configurations, and fixed coordinates
  – Optionally transform proximities with linear, ordinal, smooth ordinal, or spline functions
  – Specify multidimensional unfolding with identity, weighted Euclidean, or generalized Euclidean models
  – Specify fixed row and column coordinates to restrict the configuration
  – Specify initial configuration (classical triangle, classical Spearman, Ross-Cliff, correspondence, centroids, random starts, or custom), iteration criteria, and penalty parameters
  – Specify plots for multiple starts, initial common space, stress per dimension, final common space, space weights, individual spaces, scatterplot of fit, residuals plot, transformation plots, and Shepard plots
  – Specify output that includes the input data, multiple starts, initial common space, iteration history, fit measures, stress decomposition, final common space, space weights, individual spaces, fitted distances, and transformed proximities
  – Write common space coordinates, individual weights, distances, and transformed proximities to a file

System requirements

- Software: SPSS Base 14.0
- Other system requirements vary according to platform

SPSS

To learn more, please visit www.spss.com.
For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SCT14SPC-0705