Text Explorer – Analyzing Unstructured Text Data

The Text Explorer platform can be used to analyze unstructured text data, such as text found in comment fields in surveys or in incident reports. In JMP® Pro, additional tools for analyzing and mining unstructured text data are provided. These tools are briefly described below.

The text data must first be prepared for analysis. See the page **Text Explorer – Describing Unstructured Text Data** and the JMP Help for information on using the Text Explorer platform to prepare text data for analysis.

Analyzing Unstructured Text

Text analysis involves transforming prepared text data into a document term matrix (DTM). Each row in the DTM corresponds to a document (a cell in a JMP data table), and each column in the DTM corresponds to a term. The DTM is then used as an input in all analyses.

A number of analysis options are available from the top red triangle menu in the text explorer analysis report:

Example: Pet Survey.jmp (Help > Sample Data Library)

- **Latent Class Analysis** groups documents into clusters of similar documents.
- **Latent Semantic Analysis** is a dimension reduction technique similar to principal components analysis (the specification window with default settings is shown).
- **SVD Scatterplot Matrix** (shown to the right) graphically displays the results of **Latent Semantic Analysis**.
- **Topic Analysis** groups words into themes or topics, and is similar to factor analysis.
- **Cluster Terms** or **Cluster Documents** performs hierarchical clustering of the terms or documents.

Some of these methods reduce the dimension of the DTM by applying singular value decomposition (SVD).
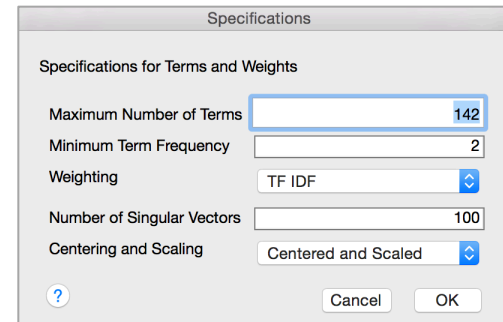
Saving Results

Text Explorer analysis results, such as the DTM, can be saved to the data table. The resulting columns can then be used as predictors in other analyses.

- To save information that corresponds to documents, save the Document Term Matrix, Document Singular Vectors or Document Topic Vectors.
- To save information that corresponds to terms, without respect to the specific documents, save the Term Table, Term Singular Vectors or Term Topic Vectors.

Note: For more details on text analysis, including information on analysis options, the DTM, singular value decomposition, and interpretation of the results, see Chapter 12 in the book *Basic Analysis* (under **Help > Books**) or search for "Text Analysis Options" in the JMP Help.