

テキストエクスプローラ - 非構造化テキストの分析

JMP PRO テキストエクスプローラは、商品のレビューコメントやソーシャルメディアの投稿、アンケート調査の自由回答、医療現場のインシデントレポートなど、構造化されてないテキストデータを分析するためのプラットフォームです。JMP Pro には非構造化テキストを分析するための手法が数多く備わっています。

テキストは予め分析用の前処理が施されている必要があります。テキストエクスプローラをテキストの前処理に用いる方法は、テキストエクスプローラ - 非構造化テキストの記述もしくは JMP のヘルプに記載されています。

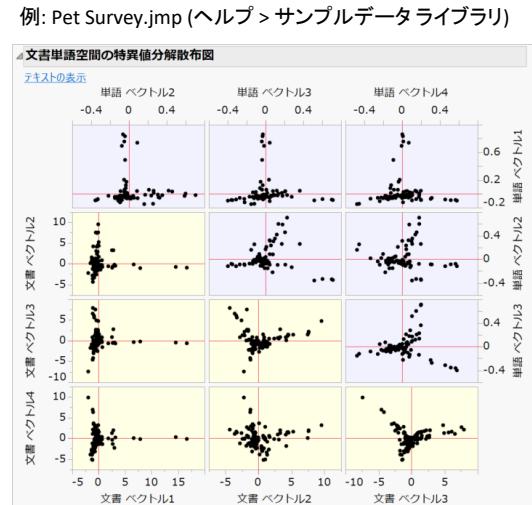
非構造化テキストの解析

前処理の済んだテキストデータを文書単語行列 (DTM; Document-Term Matrix) へ変換して、テキストを分析します。文書単語行列の各行は、1つの文書 (JMP データテーブルの 1 セル) に対応しています。文書単語行列の各列は、1 つの単語に対応しています。文書単語行列は全ての分析でインプットとして用いられます。

テキストエクスプローラのレポート内の一一番上の赤い三角ボタンから文書単語行列の保存を選択することができます。さらに、赤い三角ボタンから分析オプションを選択できます：

- **潜在クラス分析**を実行すると、類似の文書が同じクラスターに属するように分類されます。
- **潜在意味分析**は主成分分析と似た次元削減の手法です。
- **トピック分析**は因子分析に対応し、単語の負荷量の表をトピックごとに表示します。
- **クラスター 単語**もしくは**クラスター 文書**は階層型クラスター分析の結果を表示します。

ここでは**潜在意味分析**, **特異値分解**を選択して、初期設定そのまま使用します。特異値分解のレポートの赤い三角ボタンから**特異値分解 散布図行列**を選択すると、右図のようなグラフを表示できます。



結果の保存

文書単語行列などのテキストエクスプローラの分析結果をデータテーブルに保存できます。データテーブルの列は他の分析で予測変数として利用できます。

- 文書に関する情報を保存するときは、**文書単語行列の保存**、**文書 特異ベクトルの保存**もしくは**文書 トピックベクトルの保存**を選択します。
- 文書との関係を考慮せずに、単語に関する情報を保存するときは、**単語テーブルの保存**、**単語 特異ベクトルの保存**もしくは**単語 トピックベクトルの保存**を選択します。

注意: 分析オプション、文書単語行列、特異値分解及び結果の解釈などテキスト解析の詳細に関しては、JMP のヘルプか**基本的な統計分析**(ヘルプ > JMP ドキュメンテーションライブラリ以下)の 12 章で「テキストの分析オプション」と検索してご確認ください。