

# テキストエクスプローラ - 非構造化テキストの記述

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の単語や句の頻度を探索するためのプラットフォームです。結果は度数表やワードクラウドに要約されます。再コード化、句の追加、ストップワードの追加といったツールは、データのクリーニングや分析作業の絞り込みに役立ちます。

注意: JMP Pro を用いてテキストデータをさらに解析できます。詳細は[テキストエクスプローラ - 非構造化テキストの分析](#)に記載されています。

## テキストデータの要約

例: Pet Survey.jmp (ヘルプ > サンプルデータライブラリ)

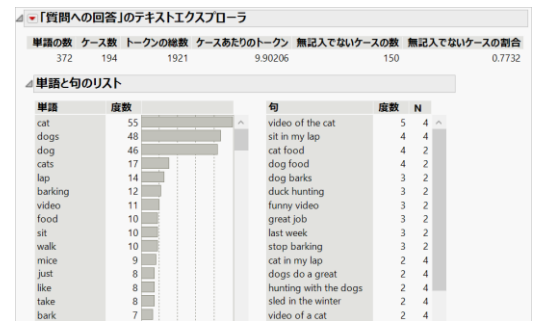
1. データテーブルから分析 > テキストエクスプローラを選択します。
2. 列の選択から質的変数または非構造化テキストの変数を選択し、テキスト列をクリックします。質的変数には赤もしくは緑の棒グラフ、非構造化テキストの変数にはグレーのライン(≡)が表示されます。
3. 言語から適切な言語を選択します。この例では英語を用います。
4. OK をクリックします。デフォルトでは JMP は以下を表示します:



- 単語とケースの数の要約。
- 単語と句のリストに再頻出の単語と句が要約されます。

## 前処理とターゲティング

語幹抽出や構文解析を含む数多くのテキストの前処理オプションを赤い三角ボタンから選択できます。さらに単語と句のリストから追加のオプションを選択できます。



- 単語もしくは句を削除する場合は、単語もしくは句を選択してから右クリックして、**ストップワードの追加**を選択します。この例では cat, cats, dog, dogs をストップワードとして選択します。
- 句を単語のリストに追加する場合は、句を選択してから右クリックして、**句の追加**を選択します。
- 単語をまとめる、もしくは値を変更する場合は、単語を選択してから右クリックして、**再コード化**を選択します。

これらは、主要テーマの重要性の発見と定量化をよりよく行うために、分析作業を絞り込むのに役立ちます。

## ワードクラウドを用いたテキストデータの可視化

前処理の済んだテキストデータのワードクラウドを作成するときは、赤い三角ボタンから**表示オプション > ワードクラウド**を選択します。

ワードクラウドをカスタマイズするオプションをワードクラウドの横の赤い三角ボタンから選択して、値に応じて単語の色を分けたりすることができます。この例では中心から配置と任意の色を用いています。



注意: テキストエクスプローラの追加情報に関しては、JMP のヘルプが[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「テキストエクスプローラ」と検索してご確認ください。