

データの加工・分析

THUMGY Dataでのテーブルのデータの加工や分析は、主にメニューバーの《加工》および《分析》からメニューコマンドを選択して実行します。

◇機能一覧

メニューコマンド名		機能	実行結果の出力先
加工	抽出	既存のテーブルからレコードまたはフィールドを抽出し、新規テーブルを作成する	新規テーブルへ出力
	結合	2つのテーブルのレコードを結び付け、新規テーブルを作成する	新規テーブルへ出力
	統合	テーブルの末尾に別のテーブルのレコードを追加し、新規テーブルを作成する	新規テーブルへ出力
	新規フィールド追加	テーブルに新しいフィールドを追加する	既存テーブルへ出力
	並べ替え	指定されたフィールドの値でレコードを並べ替え、新規テーブルを作成する	新規テーブルへ出力
	集約	集約対象フィールドのレコードを“,”区切りで一つのレコードとして出力する	新規テーブルへ出力
	補完	対象フィールドのレコードが空白の場合、直前の空白ではないレコードの値を入力する	新規テーブルへ出力
	グループ抽出	既存のテーブルからレコードまたはフィールドをグループごとに抽出し、新規テーブルを作成する	新規テーブルへ出力
	キーワード抽出	指定したキーワードが特定のフィールドに含まれている場合に指定したキーワードを出力する	新規テーブルへ出力
	文字列分解	フィールドの文字列を区切り文字で分割して縦に出力する	新規テーブルへ出力
	縦積	テーブルのレコードを抽出して縦に統合する	新規テーブルへ出力
分析	要約	対象のフィールドの値ごとに集計する	画面出力または新規テーブルへ出力
	クロス集計	縦軸と横軸となるフィールドを指定して、2つの軸の値を元にデータをグループ化し、グループごとに、数値フィールドを集計する	画面出力または新規テーブルへ出力
	グループ集計	対象フィールド（キーフィールド）の値ごとに連番を付与して値を集計する	新規テーブルへ出力
	年齢調べ	指定した「基準日」からの日数の間隔ごとに値を集計する	画面出力または新規テーブルへ出力
	偏差値	対象のフィールドの値ごとに集計し偏差値を計算する	画面出力または新規テーブルへ出力
	重複	指定したフィールドに重複した値がないか検査する	画面出力または新規テーブルへ出力
	あいまい重複	対象のフィールドのほぼ同一の値（あいまい重複）を検出する	新規テーブルへ出力

※次ページに続きます。

◇機能一覧

メニューコマンド名		機能	実行結果の出力先
分析	順番検査	指定したフィールドが順番に並んでいるか検査する	画面出力または新規テーブルへ出力
	欠落	順番に並んでいるデータの欠落を検出する	画面出力または新規テーブルへ出力
	ベンフォード分析	フィールド値の最初の桁、または、指定の桁の組み合わせの数をカウントし、実数とベンフォードの法則による予測数を比較する	画面出力または新規テーブルへ出力
	データ検証	インポートしたデータがフィールドのデータ型と一致しているか検証する	画面出力または新規テーブルへ出力
	ヒストグラム	文字または数値の値をグループ化し各グループのレコード数をカウントする	画面出力または新規テーブルへ出力
	階層化	数値フィールドのレコードを数値の間隔でグループ化して、件数のカウント・数値を小計する	画面出力または新規テーブルへ出力
	エクスポート	テーブルのデータを別のアプリケーションのファイルに書き出す	ファイル出力
	フォルダエクスポート	フォルダ内のテーブルのデータを別のアプリケーションのファイルに書き出す	ファイル出力
サンプリング	サンプリング	テーブルから一部のレコードをサンプルとして抽出する	新規テーブルへ出力
	サンプルの評価	抽出したサンプルのテスト結果から、誤謬率/額が許容できる範囲にあるかどうかを評価する	画面出力

※メニューコマンドは、順次追加されます。

◇実行結果の出力先

メニューコマンドの実行結果は、テーブルに出力されるもの、画面に出力されるもの、外部アプリケーションのファイルに出力されるものがあります。

<画面出力の例（《要約》の実行結果）>

The screenshot shows the THUMGY Data interface with a summary report for '売上データ2017上期'. Callouts provide the following information:

- タブにはメニューコマンド名が表示されます。** (The tab displays the menu command name.)
- 操作の実行日時** (Operation execution date and time: 2023/02/24 11:10:20)
- 操作の対象となったテーブル** (Table targeted for operation: 売上データ2017上期)
- 操作の内容（スクリプト形式）** (Operation content in script format: SUMMARIZE ON [営業担当者] SUBTOTAL [金額] OTHER TO SCREEN PRESORT)
- レコード件数など（コマンドによって異なります。）** (Record count, etc. (varies by command))
- 実行結果** (Execution result table):

営業担当者	合計_金額	カウント件数
2 0010	16,380	1
3 0021	434,950	7
4 0024	8,913,020	53
5 0038	36,200	2
6 0045	9,922,190	49
7 0052	10,293,430	49
合計	29,616,170	161

ドリルダウン機能
画面出力の場合で、実行結果の値がリンクになっている場合は、その値をダブルクリックすることで、関連するレコードを表示することができます。

なお、出力先に関わらず、実行日時、対象テーブル、操作の内容などの情報は、ログに記録されています。