

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）を検出します。

◇機能

《あいまい重複》は、指定したフィールドのほぼ同一の値（あいまい重複）した値がないか検査する機能です。

《あいまい重複》の結果は、テーブルに出力することも可能です。

似通ったテキストを抽出し、テキストを似せてダミーで登録されたデータがないか調べることができます。

	伝票番号	法人名	金額
1	100001	株式会社AAA	1,000
2	100002	株式会社AAB	1,000
3	100003	株式会社BBB	700
4	100003	株式会社CCC	800
5	100004	株式会社DDD	900



	伝票番号	法人名	金額
1	100001	株式会社AAA	1,000
2	100002	株式会社AAB	1,000

法人名がほぼ同一の値のレコードが抽出される

◇メニュー

《分析》-《あいまい重複》

◇設定内容

The screenshot shows the 'あいまい重複' (Fuzzy Duplicate) settings window. The interface includes the following elements:

- 1**: テーブル名 (Table Name) field, set to 'あいまい重複'.
- 2**: 相違のしきい値 (Difference Threshold) input field, set to '1'.
- 3**: あいまい重複の対象フィールド (Fuzzy Duplicate Target Fields) table, listing '伝票番号' (Invoice No.) and '法人名' (Company Name).
- 4**: 出力フィールド (Output Fields) table, listing '伝票番号' (Invoice No.), '法人名' (Company Name), and '金額' (Amount).
- 5**: Checkboxes for selecting output fields.
- 6**: フィールド選択 (Field Selection) button.
- 7**: オプション (Options) section with checkboxes for '相違のパーセント(%)' (Difference Percentage), '完全な重複を含める' (Include Complete Duplicates), and '結果サイズ(%)' (Result Size).
- 8**: IF条件 (IF Condition) section with a menu icon.
- 9**: '完全な重複を含める' (Include Complete Duplicates) checkbox.
- 10**: 保存先 (Save To) input field.
- 11**: 実行 (Execute) button.
- 12**: キャンセル (Cancel) button.

1. テーブル名 : 操作の対象となっているテーブルが表示されます。
2. 相違のしきい値 (必須)
 - : 許容される相違の文字数を指定します。
 - 値を大きくすると、あいまい度が高い値 (相互の関連性が低い文字列) が含まれるため、一致の件数が多くなります。
3. あいまい重複の対象フィールド (必須)
 - : あいまい重複検査をするフィールドを選択します。
4. 出カフィールド : フィールドの一覧から、新規テーブルに出力するフィールドをクリックで選択します。
 - フィールドは、複数選択できます。
 - 選択した順に番号が表示され、フィールドの並び順が決まります。
5. フィールドの全選択/全解除
 - : チェックボックスのオン/オフで、すべてのフィールドの選択/解除が行えます。
6. 《フィールド選択》ボタン
 - : 「フィールドの選択」画面を開き、フィールドの選択・並び順の変更を行うことができます。
7. オプション (任意) : 重複した値のグループ番号を追加します。
 - 相違のパーセント(%) : 文字列のうち異なってもよい割合を制限します。
 - 結果サイズ (%) : レコード数に対して、出力結果のセット件数の割合を制限します。
 - 完全な重複含める : あいまい重複の件数結果に完全な重複も含めて出力します。
8. IF条件 (任意) : 抽出条件を設定する場合は、「IF条件」に条件式を入力します。
9. 《式ビルダー》アイコン
 - : 「IF条件」の条件式を入力するために、式ビルダーを開いて使用できます。
10. 保存先 (必須) : 実行結果の出力先となる新規テーブル名を入力します。
 - テーブル名の長さは、64文字までです。
 - ファイルの保存先は、プロジェクトが保存されているフォルダの「04_table」フォルダです。
 - 《保存先》ボタンで、保存するフォルダを変更できます。使用可能なフォルダは、プロジェクトが保存されているフォルダの「04_table」フォルダおよびそのサブフォルダのみです。
11. 《実行》ボタン : 実行します。
12. 《キャンセル》ボタン : 実行せず画面を閉じます。

◇操作ログ

FUZZYDUP ON FIELDS [あいまい重複対象フィールド] **OTHER FIELDS** [その他出力フィールド1] [その他出力フィールド2] …または**ALLFIELDS LEVDISTANCE** 相違のしきい値 **オプション** **IF 条件式** **TO** テーブル名 **OPEN**

※オプションを選択した場合は、選択したオプションが記述されます。

オプション	ログ
相違のパーセント(%)	DIFFPCT 割合
結果サイズ (%)	RESULTSIZ 割合
完全な重複含める	EXACT

※「IF条件」を設定した場合は、IF条件式が記述されます。