

# 日付データを加工する

## ここで使う！

2015/01/01のような形式のデータから、年、月、日、曜日、平日休日等の情報を特定し、新しい列として追加します。また、特定の日付からの経過日数を計算させます。取り出した月、曜日、平日休日のデータは集計に用いたり、各種の分析で説明変数とすることもできます。

## どうする？

文字列の日付データから特定の時系列属性を取り出すには、「日付時刻処理」アイコンを使用します。

文字列としてインポートした1列目「年月日」から、「年」「月」「日」「曜日」「休日フラグ（平日か、休日か）」「休日名」を特定し、元のデータの横に並べて新しいデータとします。オリジナルのデータと、操作によりできたデータは下記のようになります。

	A	B	C	D
1	年月日	東京の平均気温	降水量の合計(mm)	
2	2014/1/1	9.6	0	
3	2014/1/2	7.3	0	
4	2014/1/3	5.9	0	
5	2014/1/4	6.5	0	
6	2014/1/5	5.4	0	
7	2014/1/6	5.3	0	
8	2014/1/7	5.5	0	
9	2014/1/8	7.3	15.5	
10	2014/1/9	7.2	4	

オリジナルのデータ

result								
	年月日	東京の平均気温	東京の降水量の合計(mm)	年	月	日	曜日	休日フラグ
1	2014/1/1	9.600	0.000	2014	1	1	水	休日
2	2014/1/2	7.300	0.000	2014	1	2	木	平日
3	2014/1/3	5.900	0.000	2014	1	3	金	平日
4	2014/1/4	6.500	0.000	2014	1	4	土	休日
5	2014/1/5	5.400	0.000	2014	1	5	日	休日
6	2014/1/6	5.300	0.000	2014	1	6	月	平日
7	2014/1/7	5.500	0.000	2014	1	7	火	平日

日付時刻処理の結果

「年」「月」「日」「曜日」「休日フラグ」などのデータが新しい列として追加されます。

データは、気象庁様のダウンロードサイトからダウンロードしたものです。  
<http://www.data.jma.go.jp/gmd/risk/obsdl/>

## 準備

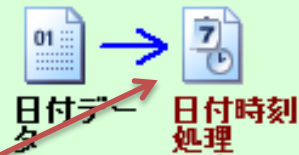
### Object Browser

- データ
- データベース
- データ操作
- クリーニング

① 画面左側のオブジェクトブラウザ（Object Browser）から、「データ操作」の「日付時刻処理」アイコンを、プロジェクトボードへドラッグ&ドロップします。

- 列属性変更
- 正規化
- 層別サンプリング
- 匿名化
- 再配置
- データハンドリング
- 日付時刻処理**
- 行選択
- 日付選択
- 集計
- パラメータ
- データエク

① ダウンロードした「日付データ」ファイルをドラッグアンドドロップでインポートします。



④ 日付時刻処理アイコンをダブルクリックして、設定画面を立ち上げます。

② データアイコンから日付時刻処理アイコンへ、マウスの中ボタンで矢印を繋ぎます。

■ 日付の基本情報を取得（年、月、曜日などを用いて分析したい場合に便利です）

手順

The screenshot shows the '日付処理' (Date Processing) dialog box with the following sections and callouts:

- 対象時刻 (Target Date/Time):**
  - ① 「列形式」から「年月日」を選びます。 (Select '年月日' from the '列形式' dropdown.)
  - 列指定: 年月日 (日付形式=自動検出)
  - 時刻指定: 2015年 1月19日
- 基本出力項目 (Basic Output Items):**
  - ② 日付フォーマットを指定する場合、「形式」をクリック、選択をします。このデータは日付形式を自動検出できます (Click '形式' to select a date format. This data can be automatically detected as a date format.)
  - 年  月  日  時  分  秒
  - 午前・午後  年通算週  月通算週  年通算日  休日フラグ  休日
  - 曜日名: 月火水木金土日,Error
- 期間計算 (Period Calculation):** オフセット日付
- 基準時刻 (Reference Date/Time):**
  - ③ 「基本出力項目」で「年」「月」「日」「休日フラグ」「休日」「曜日」など必要な情報を選択します。「曜日」は種類の選択(英語表記、数字付など)も可能です。 (Select necessary information like '年', '月', '日', '休日フラグ', '休日', '曜日' in the '基本出力項目'. '曜日' can also be selected by type (English notation, with numbers, etc.).)
  - 列指定: 年月日 (日付形式=自動検出)
  - 時刻指定: 2015年 1月19日
  - 基準時刻から  基準時刻まで
- 出力単位 (Output Unit):**
  - ④ 元のデータも必要な場合「元データ出力」にチェックします。 (Check '元データ出力' if you also need the original data.)
  - 年数  月数  週数  日数  時数  分  秒数  満年数  満月数
- 元データ出力  OK Cancel

ワンポイント  
アドバイス

◆ 基本出力で出力されるデータは？

年月日などは整数として、曜日などは文字列として出力されます。計算されたデータは分析の際には説明変数とすることができます。

■ 期間計算を行う（経過日数や経過時間を集計、分析する場合に便利です）

手順

The screenshot shows the '日付処理' (Date Processing) dialog box with the following settings and callouts:

- 対象時刻 (Target Time):**
  - 列指定: 年月日 (日付形式 = 自動検出) [形式]
  - 時刻指定: 2015年08月03日 11:35:16
- 基本出力項目 (Basic Output Items):**
  - 年,  月,  日,  時,  分,  秒
  - 午前・午後,  年通算週,  月通算週,  年通算日,  休日フラグ,  休日名
  - 曜日,  月末,  次月初
- 曜日名 (Day Name):** 月,火,水,木,金,土,日,Error
- 期間計算 (Period Calculation):** オフセット日付
- 基準時刻 (Reference Time):**
  - 列指定: 年月日 (日付形式 = 自動検出)
  - 時刻指定: 2015年08月03日 11:32:52
- 基準時刻から (From Reference Time):**  基準時刻から,  基準時刻まで
- 出力単位 (Output Unit):**
  - 年数,  月数,  週数,  日数,  時数,  分数
  - 秒数,  満年数,  満月数
- 元データ出力 (Original Data Output):**  元データ出力

Callouts (in green boxes):

- ① 「列形式」から「年月日」を選びます。
- ② 「時刻指定」で基準となる年月日を指定します。
- ⑤ 「出力単位」で「日数」など計算させたい値を選びます。
- ③ 元のデータも必要な場合「元データ出力」にチェックします。

## ワンポイント アドバイス

### ◆ 期間計算で出力されるデータは？

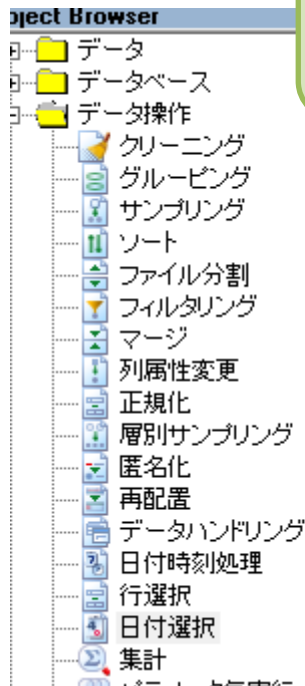
基準時刻から対象列までの日数や時間数が計算されます。購買履歴データならば、リーセンシの計算に役立ちます（RFM 分析はアイコンとして準備されており、日数計算は不要です）。Decision Tree をはじめとしたモデリングやクラスタ分析では、説明変数として用いることも可能です。

また、「グルーピング」アイコンや「集計」アイコンと組み合わせることにより、生年月日から年代を計算したり、期間中の合計や平均を出すことができます。

## ■ (応用) 日付の基本情報を使って集計や分析をする

列追加した基本情報を用いて、集計や分析を行い、傾向をデータから読み取ります。まずは、月ごとの気温と降水量の平均を求めます。

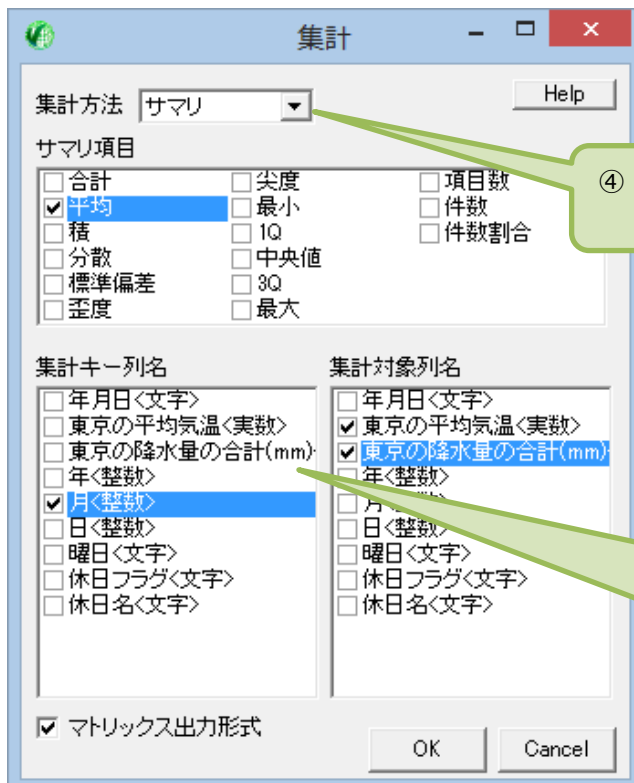
### 手順



① 画面左側のオブジェクトブラウザから、「データ操作」の「集計」アイコンを、プロジェクトボードヘドラッグ&ドロップします。

② 日付時刻処理アイコンから集計アイコンへマウスの中ボタンで矢印を繋ぎます。

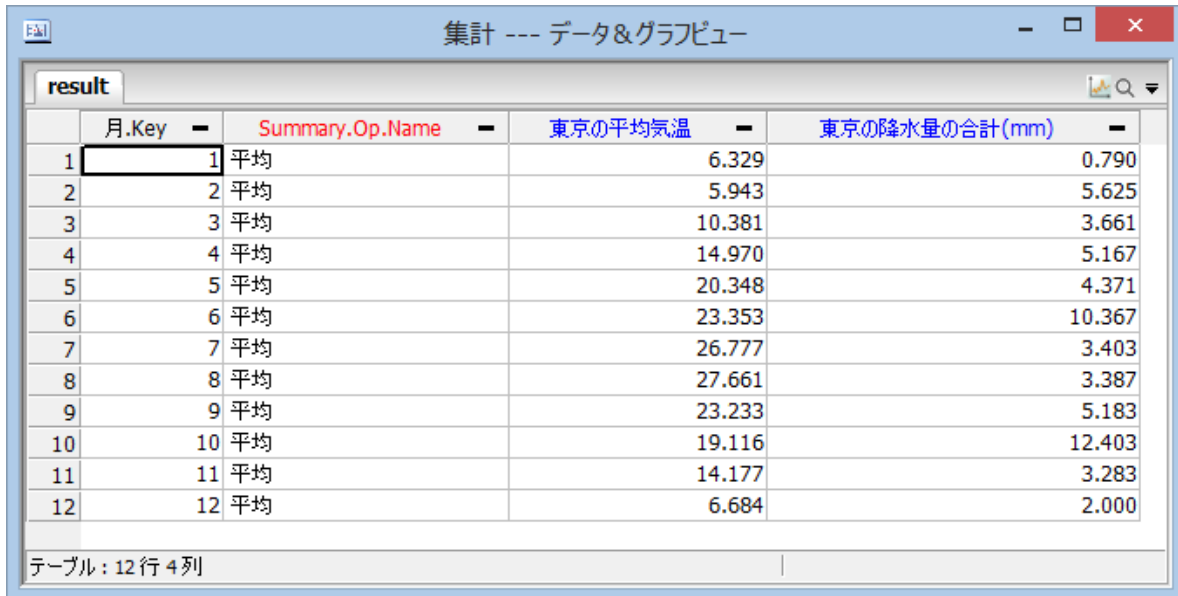
③ 集計アイコンをダブルクリックして、設定画面を立ち上げます。



④ 集計方法を「サマリ」とし、サマリ項目から「平均」をチェックします

⑤ 集計キー列名から「月」を選択し、集計対象列名から「東京の平均気温」と「東京の降水量の合計」をチェックします。月ごとの気温や降水量の平均を計算します。

集計キーは複数選択することもできます。計算結果は選択された列のクロス表での平均値です。平均気温と降水量の月ごとの推移を確認することができます。



	月.Key	Summary.Op.Name	東京の平均気温	東京の降水量の合計 (mm)
1	1	平均	6.329	0.790
2	2	平均	5.943	5.625
3	3	平均	10.381	3.661
4	4	平均	14.970	5.167
5	5	平均	20.348	4.371
6	6	平均	23.353	10.367
7	7	平均	26.777	3.403
8	8	平均	27.661	3.387
9	9	平均	23.233	5.183
10	10	平均	19.116	12.403
11	11	平均	14.177	3.283
12	12	平均	6.684	2.000

テーブル: 12行 4列