

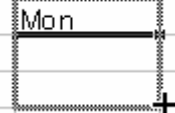


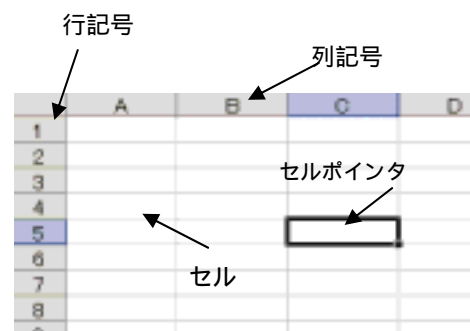
第 8 回：Excel のデータベース機能

1. Excel によるデータベースの作成

1.1. 基本用語

- Excel ワークシートの名称：行，列，セル，セル番号 (右下図)
- マウスポインタ (左下図)

	クリックするとセルポインタが移動する。ダブルクリックするとセル内容の編集ができる。
	この状態でマウスをドラッグするとセルの内容を移動したりコピーしたりできる。
	この状態でマウスをドラッグするとセルの内容をオートフィルできる (セル参照式や表示形式をコピーする)。

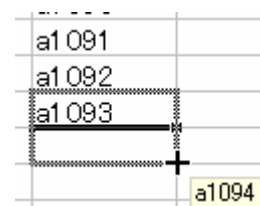


1.2. データベースとは

- 同じ規則でデータをたくさん集めたものを**データベース**という。
- Excel では，ワークシート上に作成されたデータベースを**リスト list** といい，リストは行ごとのデータ (**レコード record**) を単位とし，項目 (**フィールド field**) を列ごとに入力する。また，リストの 1 行目は見出しにする。見出しのテキストを**フィールド名**とよぶ。
- リスト作成上の注意：
 - 並べ替えをすると，元のリストの順番を復元できなくなる恐れがある。レコードの順番に何らかの意味がある場合，一意のレコード番号をつけるフィールドを用意する (複数のフィールドに分けたほうがよい場合もある)。
 - 漢字圏の言語の場合，ふりがな等を別フィールドに作成しておかないと並べ替えができない。


1.3. リストの入力のヒント

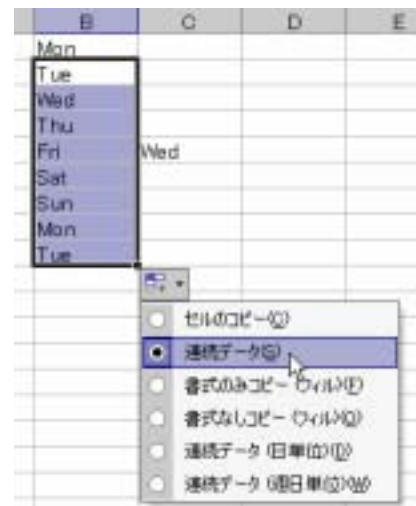
- 列方向にセルを移動： **Tab** キー；行方向にセルを移動：**Enter** キー
- 再編集：セルをダブルクリックするか，**F2** キーを押す。
- **オートコンプリート**：すでに入力されているデータを候補として提示し，入力の手間を省く (**Enter** を押して自動入力，**BackSpace** で候補をキャンセルし，通常入力を続行)。また **Alt** + **[]** キーで候補の一覧を表示できる。
 - オートコンプリートの解除は [ツール] [オプション] の「編集」タブからおこなう。
- **オートフィル**：各フィールドに設定された書式や計算式を隣接したセルにコピーする。
- **フォーム**：[データ] [フォーム] ... 入力のほか，簡単な検索に使える。
- データのコピー：書式だけ，計算式ではなく計算した結果だけ，など，コピーしたい内容はさまざま。 [編集] [形



式を選択して貼り付け] で詳細に貼り付ける内容を選べる (すぐ下の 1.4.も参照)。

1.4. セルの書式設定

- Excel の書式設定の基本：セルを選択し，[書式] [セル] (右図)
- セルの表示形式
 - 「標準」と数値，文字列，日付など
 - データの配置，小数点や記号などの表示
 - 小数点
- 文字列と数値 
 - 文字として入力
 - 0 による桁そろえ (文字列として入力，先頭に ' (シングルクオート) を入力)
 - 連続データの作成
 - ユーザ設定リスト ([ツール] [オプション])：ユーザ設定リストに登録すると新しい連続データのパターンを利用できる。
 - オートフィルオプションによるコピー方法の設定 (オートフィルオプション，右クリックしながらオートフィルすると起動，右下図)
- セルのコピーと貼り付け (左下図)



- データではなく，書式だけをコピー
 - 「書式のコピー / 貼り付け」ボタン
 - セルをコピー後，[編集] [形式を選択して貼り付け]

2. データの検索と集計

大量のデータからのパターンを抽出するデータ分析方法を「**データマイニング**」 Data Mining とよぶ。データの検索と集計は，データマイニングの基本手法である。Excel で作成したリストを検索し，リストをもとにさらに集計をとる方法を紹介する。

2.1. 不要な行・列の折りたたみ：行・列を選んで [書式] [行]または[列] [表示しない]

- 余計なものを隠すことで，スッキリとしたリストにする。
- 右クリックでも OK。

2.2. 並べ替え sort：[データ] [並べ替え]

- データベースでは，並べ替えはレコード単位でおこなう。
- レコード中のフィールドを並べ替えの基準 (**キー**) とし，各キーの順番を「昇順」「降順」のいずれかで並べ替えることができる。
- 優先順位をつけた複数のキーで並べ替えをすることも可能。

- 並べ替えの注意点：
 - Excel の並べ替え機能では、並び順は文字コード順を基本とし、かなは「五十音順」で自動的に Excel が判断して決められる (かな・カナは同一視され、カタカナの全角・半角の区別はされない)。
 - 英語以外の言語の文字順での並べ替えは Excel ではできない。(Word2000 以降には、表の機能として、各言語の並び順による並べ替え機能がある。)

2.3. 絞り込み検索 drill-down search : [データ] [フィルタ] [オートフィルタ]

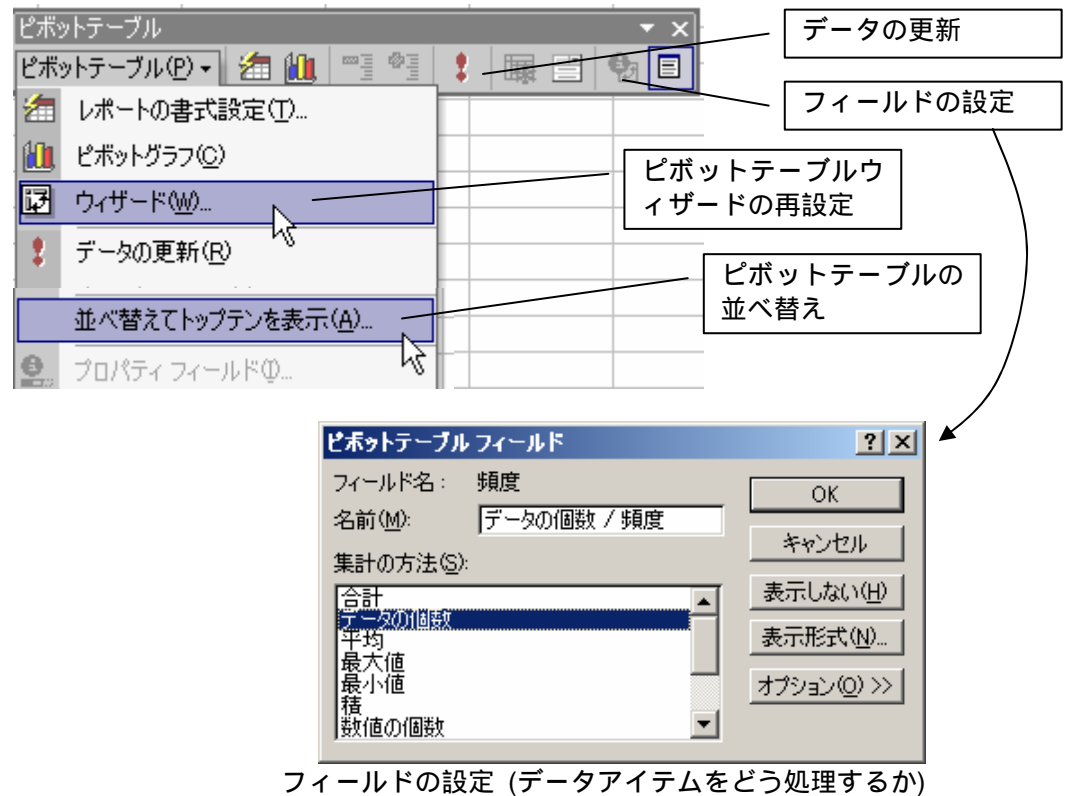
- オートフィルタの起動：リスト中にセルポインタを置き、オートフィルタを起動する。フィールド名の箇所にドリルダウン用のメニューができる。
 - フィールド名は全てのフィールドについて入力しておく。リストの見出し中に何も書かれていないフィールドがあると、オートフィルタが起動しない。
- オートフィルタの解除：[データ] [フィルタ] [オートフィルタ] のチェックをはずすとオートフィルタが解除され、通常のリストに戻る。
- 検索条件の指定：フィールド名のドロップダウンリストから、「トップテン」「オプション」「オプション」を選ぶことができる
 - 「オプション」を選ぶと、最大 2 つの条件について、AND/OR 検索や細かい抽出条件の指定ができる。(3 つ以上の条件を使うには、
 - 「オプション」での抽出条件にはワイルドカード (*, ?) が利用可能。* は任意の文字列 (長さは自由)、? は任意の 1 文字を表す。正規表現は使えない。
 - ドロップダウンリストから「全て」を選択すると絞り込みを解除できる。
- 【補足】絞り込んだレコードのみをコピー：Excel2002 以前の Excel では、コピーしたい範囲をマウス等で選択し、Alt + ; (セミコロン) を押して絞り込みで見えていないレコードの選択を解除してからコピーをおこなう必要があるので注意。

2.4. 条件に合うセルの色を変える：範囲を選んで [書式] [条件つき書式]

- リストの必要な範囲を選択してから [条件つき書式] を設定する。
- 最大 3 つまで条件をつけて書式を設定できる (追加は「追加」ボタンで)。

2.5. 集計 tabulation : [データ] [ピボットテーブルとピボットグラフレポート]

- 集計とは、データの個数を数えること。
- 特定の分類基準 (カテゴリー) に従ってデータを分けて数える場合、分類のカテゴリーが 2 つある場合をクロス集計 cross tabulation と呼ぶ。
- Excel では、クロス集計表 cross table はもちろん、カテゴリーが 3 つ以上の多次元集計表を作成できるほか、複数種類のデータを同時に集計することもできる。
 - 集計元のリストには、
 - 行のフィールドと集計対象とするデータ (データアイテム) を同一にすると、単純にデータの個数を数えてくれる。
 - カテゴリーは、縦横に並べるほかに、縦に複数並べることもできる。横に長い表を避けたい場合には後者のほうが都合がよい。
- グレーのフィールドボタンをドラッグしてフィールドを削除したり、移動したりできる。
 - 複数データを集計している場合、一方のデータのみを削除するには、ボタンをクリックしてチェックをはずす。
- 集計結果のセルをダブルクリックすると、新しいワークシートに集計元のレコードが抽出される。
- 「ピボットテーブル」ツールバーの主な機能：



実習データの出典

- セル.xls : セル入力の練習用
- 大相撲.xls : 平成 15 年 11 月場所の大相撲番付 (データベース機能の実習用)
- 試験サンプル.xls : 架空の試験データサンプル (集計の実習用)

- BNC.xls : 名古屋大学滝沢先生に提供いただいた British National Corpus からの用例
- 計量国語学会バックナンバーデータベース.csv : 学会ホームページよりダウンロード可能 (<http://wwwsoc.nii.ac.jp/math-ling/>)。
- 計量日本語学集成.txt : 目次は荻野綱男先生のホームページよりダウンロード可能 (<http://www.comp.metro-u.ac.jp/~oginotsu/>)。文献自体は論説資料保存会より CD-ROM にて配布されている (<http://www.ronsetsu.co.jp/keiryo/>)。