

データを入力する

- Excel で作業する際にはデータをなんらかの方法で入力しなければならない
- Excel 形式でデータが入っていれば別に問題はないが、Web ページに載っている表のデータを使いたいことはよくある
- データ入力の大原則は「できる限り手間を掛けない」。データを打ち込む作業自体にはなんの価値もない。コンピュータが読めるデータをいちいち印刷して手で再入力するのは人生の無駄遣い、給料泥棒

以下では基本的に縦横表になっているデータを前提とし、だらだらと長い文書のようなものは相手にしない。

1. ファイル経由でデータを読む

Excel はかなりいろいろな形式のファイルを扱うことができる。[ファイル]メニューの[ファイルを開く]で選択可能なファイルの種類は

1. Excel ファイル
2. HTML ファイル
3. テキストファイル
4. CSV(comma separated value)ファイル
5. その他沢山

と非常に豊富である。1 の Excel ファイルは特に説明する必要はないので略。2. の HTML ファイルは、Web からデータを手取りする場合に非常に有効な手段であり、扱いを知っているとかなり便利。3 のテキストファイルと 4 の CSV ファイルはアプリケーションや OS にあまり依存しないため、データベースを購入するところの形式になっていることが多い。

1.1 HTML ファイル

HTML ファイルは、Web の記述に利用されるデータ形式である。通常表組みは

```
<TABLE>
```

```
<TR>
```

```
<TD></TD><TD></TD>
```

....

</TR>

<TR>

...

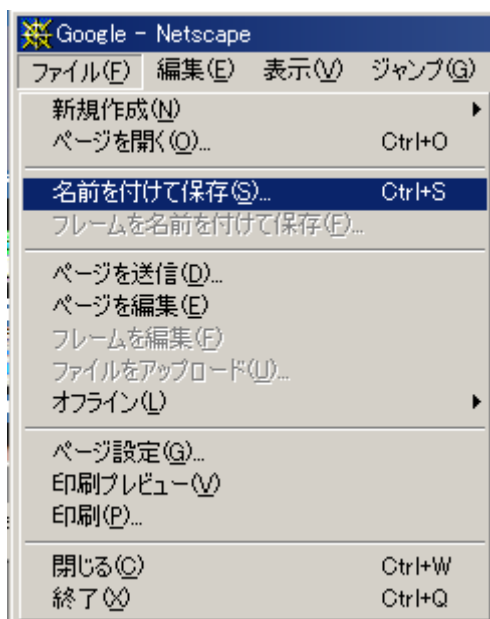
</TR>

</TABLE>

といった感じで記述されている。ただし、Excel はこの形式を理解できるのでこのことを知らなくてもあまり問題にはならない。Excel に HTML ファイルを読み込んだ場合、表組以外の部分も全部セルデータとして取り込んでしまうので、不要な部分はあとで削除する必要がある。なお、以下の操作はかなりブラウザ依存の話なので、利用しているブラウザのバージョンによっては異なる可能性も高い。

1.1.1 普通のページの保存

Web ブラウザから現在見ているページを HTML ファイルに保存する場合は、



のように、[ファイル]メニューから[名前を付けて保存]を選択する。上の図は Netscape Communicator の例だが、Internet Explorer でも同じメニュー構造¹となっている。ここで保存した HTML を Excel から読み込むには、Excel の[ファイル]-[開く]で、「ファイル

¹ [ファイル]-[名前をつけて保存]は Windows の基本的なメニュー構造なので、よっぽど偏屈なソフトでなければこれと同等の機能が同じようなメニューに用意してあるはず

の種類」を「Web ページ」にすればよい。

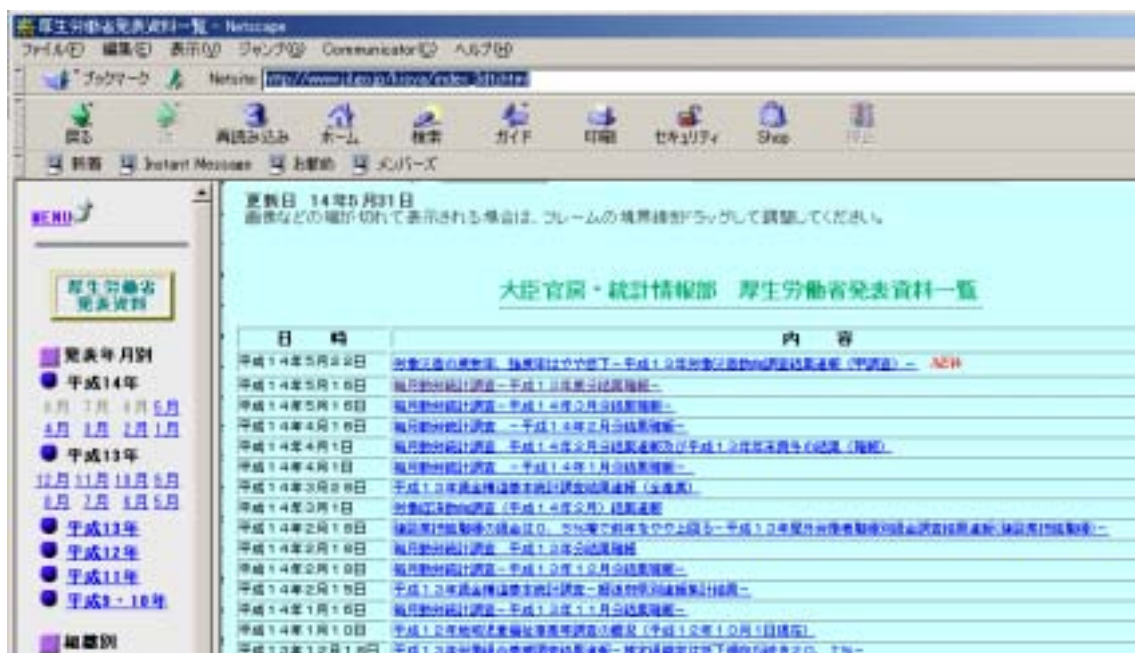
1.1.2 フレーム構造のページの保存(Netscape Communicator の場合)

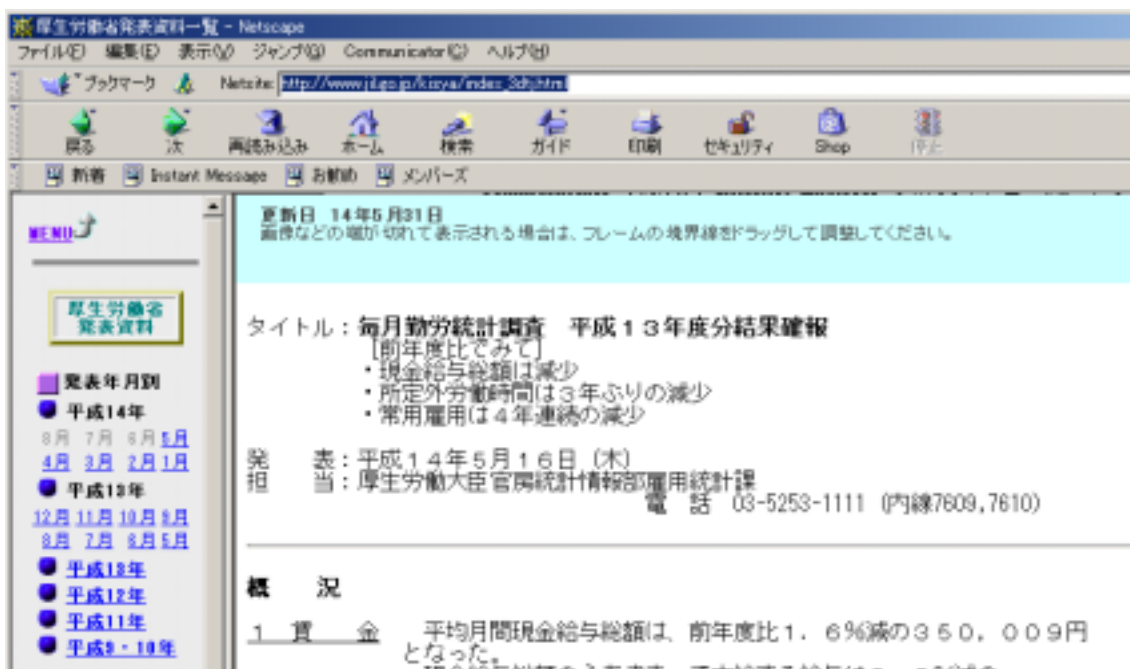
もし保存したいページがフレームを使っている場合は、ちょっとした小細工が必要になる。

例えば、厚生労働省が発表している「毎月勤労統計調査」のデータは概略が発表資料として Web に掲載されており、http://www.jil.go.jp/kisya/index_3dtj.html が統計情報部の発表資料の目次ページである。このページを開くと、下図のようなページ構成になっている。よく見ると、左側のリンク項目と右側の発表資料リストに分割されており、左右別個のスクロールバーもついている。このようなページは「フレーム構造を持っている」ページである。

さて、この場合[ファイル]-[名前をつけて保存]で保存されるのはどのような情報になるだろうか。上記 URL から「毎月勤労統計調査の平成 13 年度確報」へのリンクを開くと、次の図のようになる。上の図と左側のフレームは変わっていないことに注意。この状態で[ファイル]-[名前をつけて保存]を選択して保存を実行すると、実はどのフレームがアクティブになっているか（直前にクリックされたフレームがアクティブ）で保存されるデータの内容が異なる。

左側のフレームをクリックしてから保存を実行すると左側のフレームの内容が保存され、右側のフレームをクリックしてから保存を実行すると右側のフレームの内容が保存されるのである。これを忘れていると、データを保存したつもりでセッセとメニューを保存していることになりかねないので注意。





1.1.3 フレーム構造を持っているページの保存(Internet Explorer 6 の場合)

IE6 だと、フレーム構造を持ったページを保存しようとするするとデフォルトではフォルダを作ってフレーム構造を維持したまま全部保存しようとする。しかしながら、IE を使っている場合は後述の copy & paste でデータを Excel に貼り付ける方が楽なので、ファイルを保存してから読み込む必要は実用的にはない。

例題 1

http://www.jil.go.jp/kisya/index_3dtj.html から「毎月勤労統計調査 - 平成 13 年度分結果確報 --」を開き、その中の時系列表第 1 表を HTML ファイルとして保存し、Excel に読み込んでみる。

まず、時系列表第 1 表を Web ブラウザで表示すると、下図のようになる。

更新日 14年5月31日
画像などの個々が別で表示される場合は、フレームの種類選択をクリックして調整してください。

トップページ

時系列表第1表 賃金指数
(事業所規模5人以上) (平成12年平均=100)

年月	調査産業計		製造業		卸売・小売業		サービス業			
	規模30人以上		規模30人以上		前年度比		前年度比			
	前年度比	%	前年度比	%	前年度比	%	前年度比	%		
現金給与総額										
平成元年度	-	-	86.5	4.2	-	-	82.6	5.7	-	-
2	91.4	-	90.4	4.8	87.7	-	86.8	5.1	95.7	-
3	95.4	4.4	93.5	3.3	90.8	8.5	89.5	3.2	101.1	5.7
4	96.7	1.3	94.7	1.4	91.6	1.0	90.0	0.6	102.3	1.1
5	97.2	0.6	95.6	0.9	92.7	0.7	90.7	0.8	101.1	-1.1
6	98.7	1.5	97.4	1.9	94.3	2.2	92.8	2.3	101.8	0.7
7	99.6	0.9	98.8	1.4	96.6	2.5	95.4	2.8	102.7	0.9
8	101.2	1.6	101.0	2.3	99.2	2.8	98.4	3.2	104.0	1.3
9	102.1	0.9	102.1	1.1	101.0	1.7	100.2	1.8	104.8	0.7
10	100.4	-1.7	100.5	-1.6	99.8	-1.1	99.0	-1.1	102.1	-2.4
11	99.6	-0.8	99.8	-0.7	99.0	-0.8	98.6	-0.5	100.3	-1.9
12	100.1	0.5	100.2	0.4	100.1	1.1	100.1	1.8	100.0	-0.2

このページはフレーム構造となっているので、データが入っている右側のフレームを1回クリックしてから[ファイル]-[名前をつけて保存]で "毎勤 H13 確報.html" という名前をつけて保存する。

次に Excel から保存した HTML ファイルを開く。このとき、「ファイルの種類」を自分で変えないと HTML ファイルは候補として表示されないので注意。うまく読み込めれば、下図のような感じで読める。HTML ファイル全体を読み込んでいるので、他のページへのリンクなどが含まれている場合はそこも読み込まれている点に注意。読み込んだデータはごく普通の Excel データとして計算などが可能である。

Microsoft Excel - savetest3.xls

MS Pゴシック 11

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H) Acrobat 印刷(Pr) 印刷プレビュー(Pv) 印刷範囲設定(S) 印刷範囲の表示

G3

1
2
3
4
5
6 時系列表第1表 賃金指数
7
8 (事業所規模5人以上) (平成12年平均=100)
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

年月	調査産業計		製造業		卸売・小売業		サービス業					
	前年度比	前年度比	前年度比	前年度比	前年度比	前年度比	前年度比					
平成元年度	-	-	86.5	4.2	-	-	82.6	5.7	-	-	-	-
2	91.4	-	90.4	4.6	87.7	-	86.8	5.1	95.7	-	91.8	-
3	95.4	4.4	93.5	3.3	90.8	3.5	89.5	3.2	101.1	5.7	94.9	3.4
4	96.7	1.3	94.7	1.4	91.6	1	90	0.6	102.3	1.1	96	1.2
5	97.2	0.6	95.6	0.9	92.2	0.7	90.7	0.8	101.1	-1.1	96.3	0.3
6	98.7	1.5	97.4	1.9	94.3	2.2	92.8	2.3	101.8	0.7	98.1	1.9
7	99.6	0.9	98.8	1.4	96.6	2.5	95.4	2.8	102.7	0.9	98.5	0.4
8	101.2	1.6	101	2.3	99.2	2.8	98.4	3.2	104	1.3	99.5	1
9	102.1	0.9	102.1	1.1	101	1.7	100.2	1.8	104.8	0.7	101	1.6
10	100.4	-1.7	100.5	-1.6	99.8	-1.1	99	-1.1	102.1	-2.4	100.3	-0.7
11	99.6	-0.8	99.8	-0.7	99	-0.8	98.6	-0.5	100.3	-1.9	100	-0.3
12	100.1	0.5	100.2	0.4	100.1	1.1	100.1	1.5	100	-0.3	100.2	0.2
13	98.5	-1.6	99.5	-0.7	99.1	-1	99.7	-0.4	97.5	-2.5	99.5	-0.7
13年II期	99.9	-0.4	102.1	0.2	97.6	0.6	98.8	1	95.3	-0.9	104.2	-0.6
III	96.3	-1.1	96	0	101.5	-0.8	101.2	-0.2	101.1	-2	91.4	0.2

練習問題 1

厚生労働省職業安定局雇用政策課が平成 14 年 5 月 31 日に公表した「一般職業紹介状況（平成 14 年 4 月分）について」に含まれている季節調整済み有効求人倍率のデータを Excel に読み込み、さらに unemployment.xls に入っている失業率のデータと組み合わせ、散布図を作成しなさい。この図は、失業率（職がない人の率）と有効求人倍率（人を捜している率）を組み合わせたものであるから、一種の U-V 曲線となるはずである。

1.2. テキストファイル

テキストファイルは、たとえばメモ帳などで作成したファイルである。基本的に構造を持っていないため、Excel など扱えるようにするときにはおおまかに分けて(1)項目間を特定の文字を使って区切る（区切り文字方式）と、(2)各項目の文字数を固定長にしておき、

桁数で項目を指定する（固定幅方式）の2つの方法でデータに構造を与えることが多い。

1.2.1 区切り文字方式

改行で行の区切りを表し、区切り文字(タブ、空白、カンマなど)で項目間の区切りを表す。そのため、行の頭から順番に見ていかないとどの項目になるかは確定しない。このような方式は `unix` や `windows` では非常にポピュラーであるから、他のアプリケーションとのデータ交換の `last resort` として使える可能性が高い。

例: 空白を区切り文字にしているケース

きつね食品 2000 4356 11111

狸物産 14500 45560 1245

じゅげむじゅげむエージェンシー 12566 13456 33333

区切り文字方式の場合、注意すべき点は

(a)複数の区切り文字が連続した場合の扱い(欠損値があるとして扱うのか、連続する区切り文字は1つとして扱うのか)

(b)見た目は区別のつきにくい区切り文字(タブとスペースなど)の扱い

(c)区切ったあとのデータ形式。特に日付として解釈されてしまうケースは多い

などである。Excel でテキストデータを読み込む場合は、これらをすべて指定可能である。

1.2.2 固定幅方式

データ作成時に決めた幅に従ってデータを出力してあり、定義された幅に従ってデータを読み込んでいく形式。COBOL がこのような形式でデータを出力するのが得意であるため、メインフレームを使って作成したデータを買った場合にはこのような形式になっていることが多い。この形式の場合は、データ自体を見てもどこが区切りか分からないのが普通なので、レイアウトシートを見て「どのように区切るか」をチェックする必要がある。

この形式のデータは Excel がちょっと苦手とする形式でもある。通常の場合、レイアウトシートは(開始桁数, データ長)のペアで記述されているが、Excel ではこのようなレイ

アウトをそのまま使えず、対話的に区切り位置を指定していくからである。1行に含まれているデータの個数が少なければいいが、この手のデータは往々にして1行が数百バイトあたりするのでまちがえずに区切り位置をすべて指定するにはかなりの注意力を要する。

例: 固定幅でデータが出力されているケース

会社名: 30 カラム

項目 1: 5 カラム

項目 2: 5 カラム

項目 3: 5 カラム

012345678901234567890123456789012345678901234567890123456789

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

きつね食品 2000 435611111

狸物産 1450045560 1245

じゅげむじゅげむエージェンシー125661345633333

固定幅方式の際注意すべき点は

(a)区切り位置を数え間違えないこと。Excel の場合、区切り位置のリストを数値で入力するのではなくいちいちマウスで設定していくため1桁間違えるのは簡単である

(b)データによっては、複数行にわたるデータが1レコード分のこともある。データ定義をきちんとチェックすること

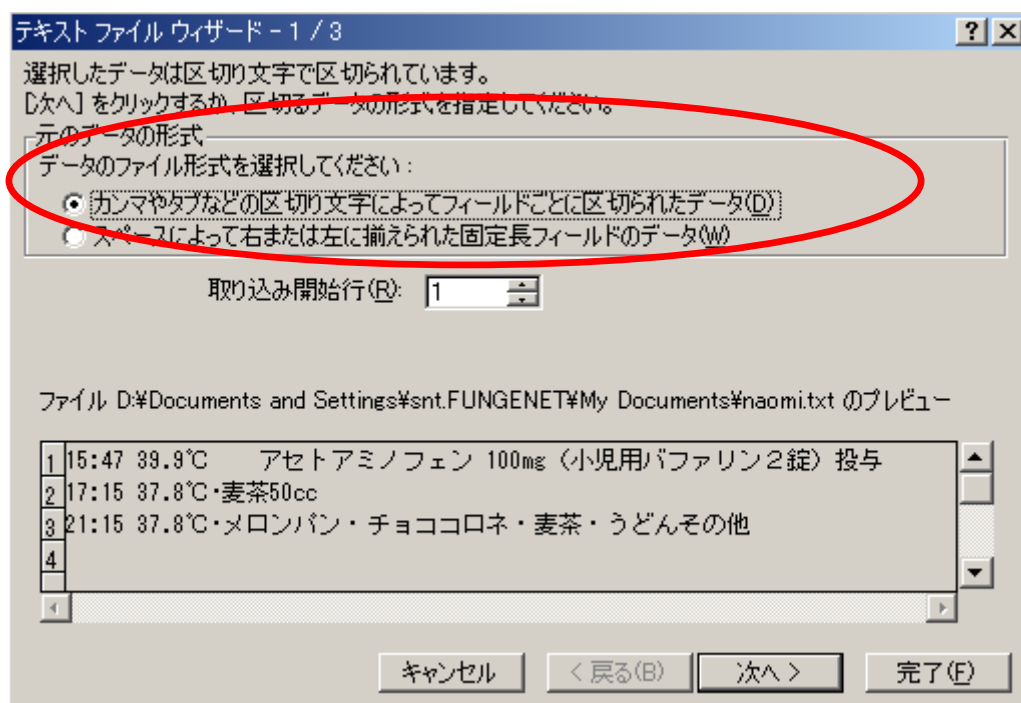
(c)区切ったあとのデータ形式の指定をしっかりとすること。手を抜くと日付に変換されてしまって救いがなくなることがある

1.2.3 テキストファイルの開き方

Excel では上記の2形式のいずれも同じ手順で開き、「区切り位置設定ウィザード」でデータの切り方を設定する。

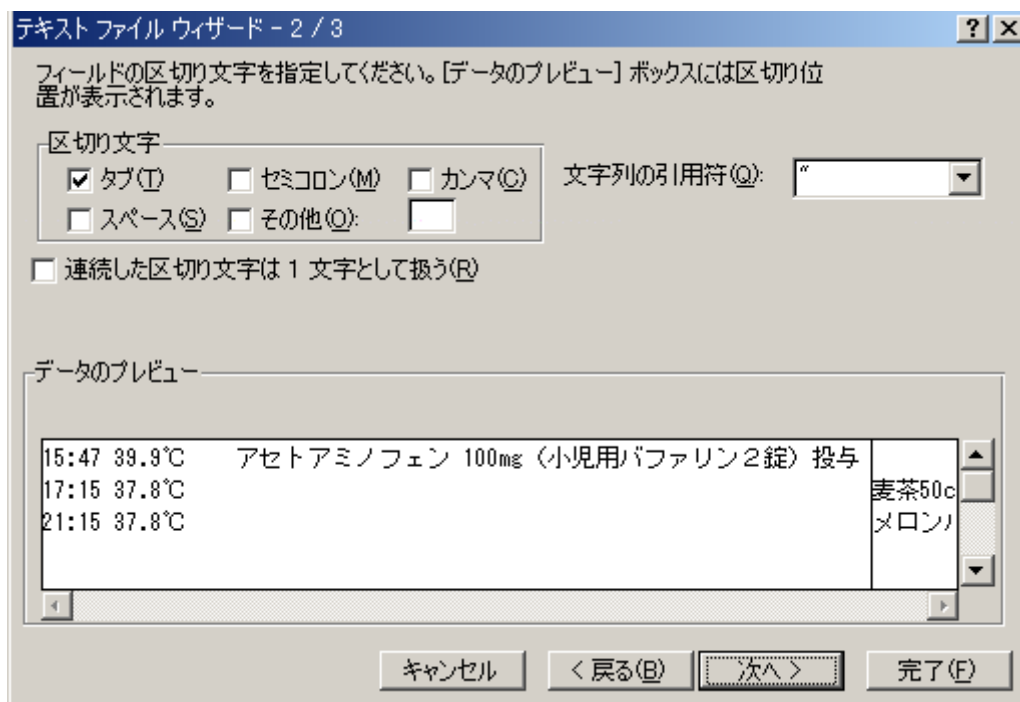
1. Excel でテキストファイルを開く場合は、[ファイル(F)]-[開く(O)]で出てくるファイルダイアログを使ってファイル名をExcelのファイルと同じように指定すればよい。

2. ダイアログの下の方にある「ファイルの種類」という項目で「テキストファイル」を選ぶと、拡張子が.prm, .txt, .csv のファイルだけが表示されるので、もし読みたいファイルの拡張子が異なる場合は「すべてのファイル」を選ぶこと。
3. Excel はファイルの拡張子を使ってデータ形式の予想をする訳ではないので、拡張子が上の 3 種類になっていなくても一向に構わない。極端な例では、テキストファイルの拡張子に .xls と付けたとしても Excel は正しくテキストファイルとして処理してくれる
4. 開いたファイルがテキストファイルであると Excel が認識した場合は、テキストファイルを読み込むための設定を聞くダイアログが開く。

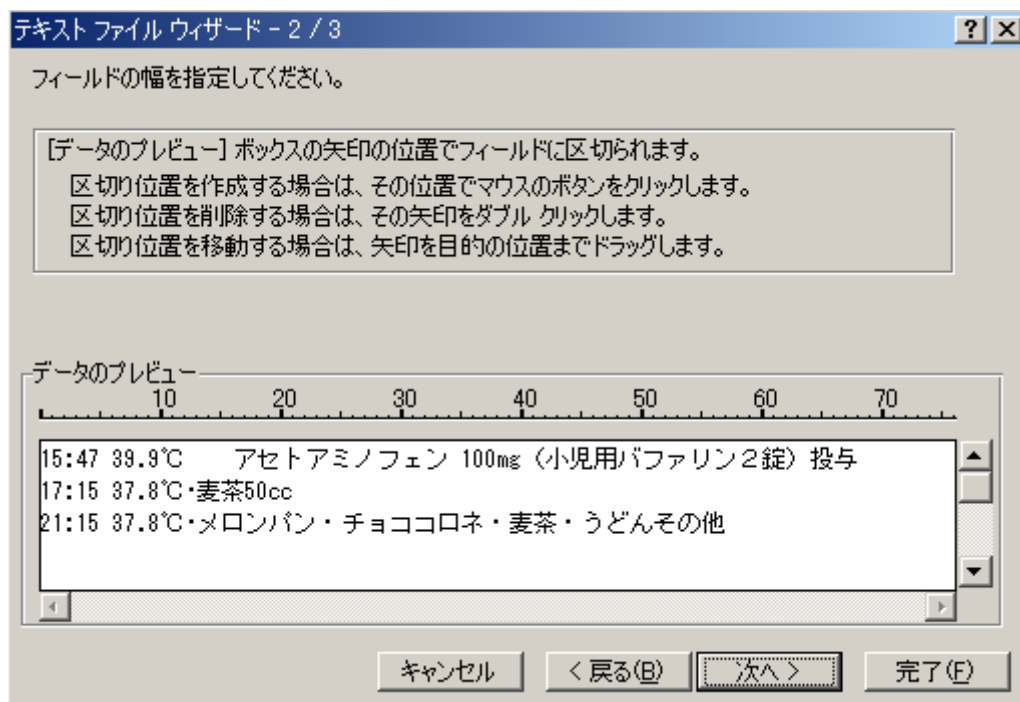


まず最初に指定するのは、データ形式が区切り文字方式なのか固定幅方式なのかである。また、取り込み開始行の指定によって、ファイルを途中から読むことも可能である。

5. もし区切り文字方式を選択した場合は、次のダイアログで区切り文字を指定できる。デフォルトではタブが指定されているが、必要に応じてセミコロン、カンマ、スペース、その他任意の文字が指定可能である。後述するように、一旦読み込んだあとで再度切り直すことも可能であるから、複雑な区切り文字が入れ子になっているようなケースでは、無理にここで全部分解しようとしなない方がよい。

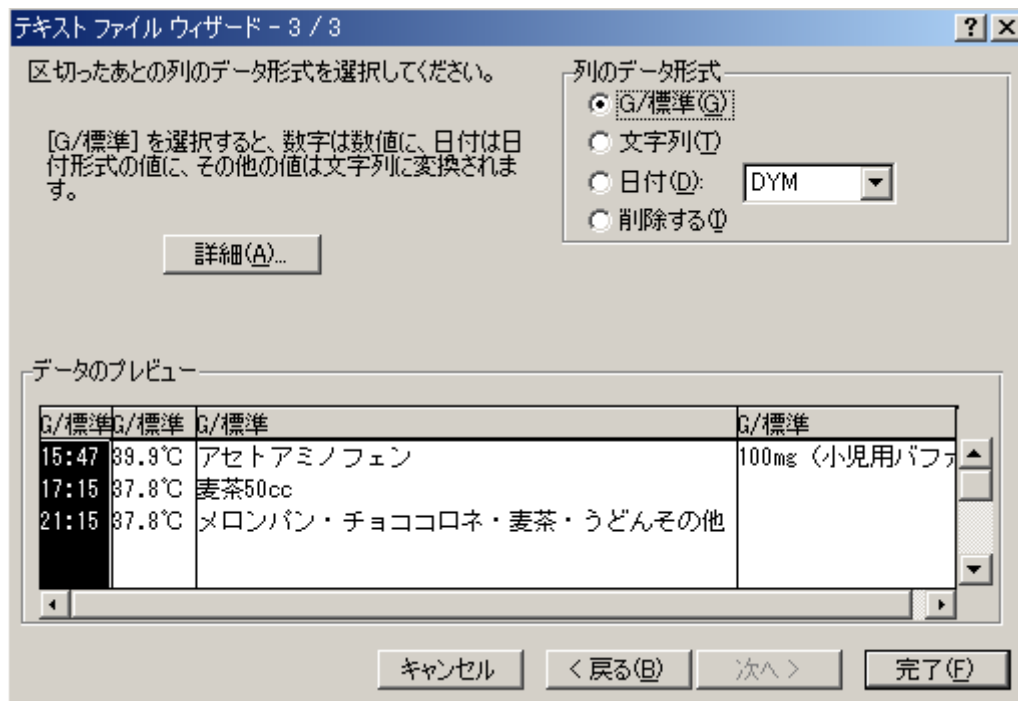


6. もし、固定幅形式を指定した場合には、マウスで区切り位置を指定していくことになる。数項目なら構わないが、固定幅形式のデータは大型コンピュータで作ったデータが多いので、1行に百項目以上のデータが入っていることも稀ではない。こうなるとこの方式はかなり辛く、またミスがどうしても避けられない。



7. 最後に、各項目の形式を指定して完了。「標準」のままにしておくと、2-3 のような文

字列が日付扱いされてしまうため注意が必要である。変換されては困る場合は、積極的にその項目を「文字列」にすること。



練習問題 2

kitsune-tab.txt と kitsune-fixwidth.txt を Excel で読み込んでみなさい。ただし、ファイルは web 上に置いてあるので Web Browser で開いた後で、[ファイル(F)]-[名前を付けて保存]でローカルファイルに保存してから作業しなさい。

kitune-tab.txt はタブ区切りのテキストファイル

kitune-fixwidth.txt は固定幅形式ファイルで、各フィールドの定義は以下の通り

項目名	開始カラム	長さ
暦年	1	4
きつね食品売上	5	3
稲荷フーズ売上	8	3
鳥居製麺売上	11	3
たぬき物産売上	14	3

1.3. CSV ファイル

- CSV ファイルは、項目間はカンマ(,)で区切った形式
- 文字列は””で囲むので、カンマを含む文字列でも扱える
- 数値は “”をつけない
- Excel では何も考えなくても CSV ファイルを読み込むことができる。ただし非常に読み込みが遅いので、遅いコンピュータを使っている際は忍耐が必要

CSV ファイルの例

“暦年”,”きつね食品”,”稻荷フーズ”,”鳥居製麺”,”たぬき物産”

1950,250,250,250,250

1955,292,308,285,313

1960,324,339,315,317

1.4. Excel ファイルその他

- 何も考えなくても Excel が勝手に読み込んでくれる
- Excel ファイルを開く場合は、マクロウィルスに注意する。特に Web から持ってくる場合は必ずチェックすること。役所がウィルス付きの Excel ファイルをバラまいた事例は何度もあるので、出所が怪しくないファイルだからといって安全である保障はない
- Web から Excel ファイルを取ってくる場合は、リンクを左ボタンでクリックするのではなく、右ボタンででてくるメニューから「リンクを名前をつけて保存」を使って必ず1度ローカルファイルに保存してからウィルスチェッカをかけること。

練習問題 3

総務省統計局の Web は多くのデータを Excel 形式で提供している。2000 年の国勢調査の集計表から全国の世帯と人口のデータを探して、都道府県ごとの 1995 年調査との人口増減率がわかりやすいグラフあるいは表を作成しなさい。

2. クリップボード経由での貼り付け

- Excel は Excel 内部でデータの貼り付けができるだけでなく、他のアプリケーションからもデータを貼り付けることができる
- しかしながら、Excel にどのような形式でデータを渡すかは元のアプリケーション次第なので、同じ操作をしても結果が同じになるとは限らない。このあたりがファイル経由と違うところ

2.1 HTML の表データの貼り付け

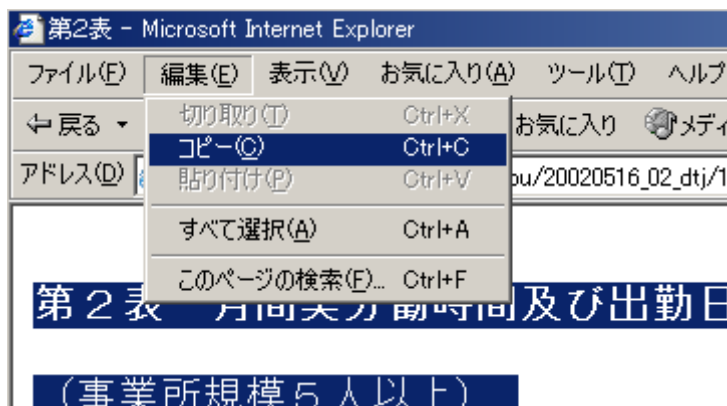
Web の表を簡単に Copy & Paste できると便利なのだが、Netscape Communicator ではこれはうまくいかないのやらしない方がよい。Internet Explorer の場合はかなりうまくいくので、以下には IE6 を使った場合の画面を提示する。

1. コピーするときは、元のアプリケーション内でコピーしたい部分をドラッグで選択し、

第2表 月間実労働時間及び出勤日数
(事業所規模8人以上)

産業	総実労働時間数		所定内労働時間数		所定外労働時間数		出勤日数	
	時間	前年度比	時間	前年度比	時間	前年度比	日	前年度比
調査産業計	152.6	-0.8	143.3	-0.5	9.3	-5.5	19.8	-0.1
鉱業	171.2	-0.8	160.2	-0.1	11.0	-10.6	21.4	0.0
建設業	168.8	-0.5	158.9	-0.3	10.0	-6.4	21.1	0.0
製造業	161.6	-1.3	149.3	-0.3	12.3	-10.5	19.9	0.0
電気・ガス業	154.2	0.1	143.8	-0.1	10.4	2.7	19.0	0.0
運輸・通信業	168.3	-0.6	149.6	-0.2	18.7	-4.1	20.3	-0.1
卸売・小売業	139.1	-1.4	133.9	-1.4	5.2	-1.3	19.7	-0.1
金融・保険業	150.2	0.2	141.5	0.2	8.7	1.7	19.3	0.0
不動産業	156.6	-0.5	148.6	-0.8	7.0	5.7	20.3	-0.1
サービス業	147.9	-0.3	140.1	-0.1	7.8	-2.7	19.5	-0.1

2. [編集]-[コピー]を使って選択範囲をコピーする



3. Excel で貼り付けたいセルをアクティブにしてから、Excel 内部での貼り付けと同様に貼り付ける。

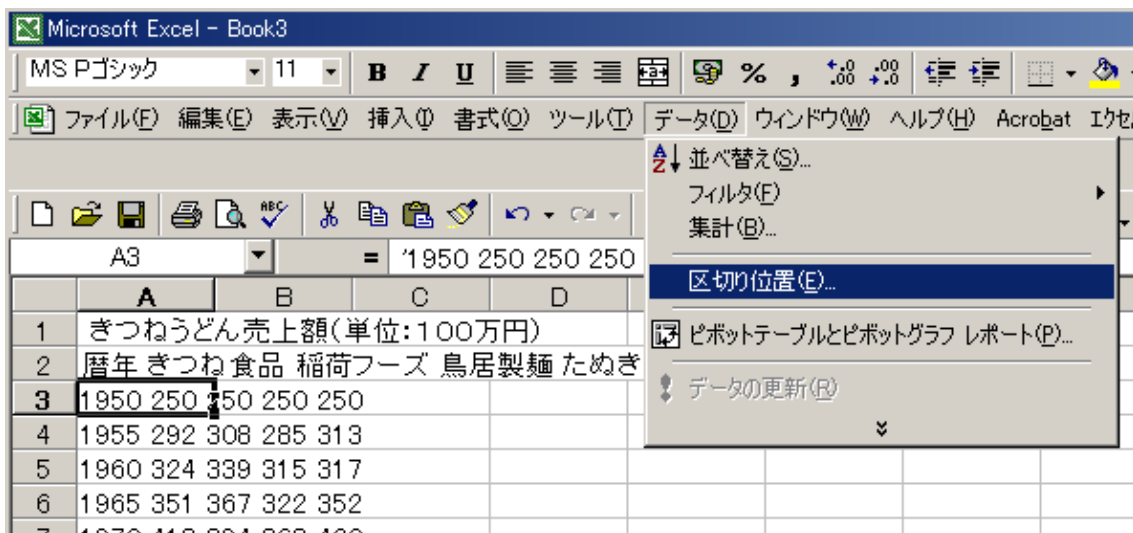
2.2 テキストデータの貼り付け

テキストデータを貼り付ける場合も、元々のデータが入っているアプリケーションによってどのような形式でデータが貼り付けられるかはよくわからない。しかし、最低でもそのデータの1行が1つのセルに入る形式では入力される。

The screenshot shows Microsoft Excel with a table of sales data. The active cell is D6. The table content is as follows:

	A	B	C	D	E
1	ぎつねうどん売上額(単位:100万円)				
2	暦年 ぎつね食品 稲荷フーズ 鳥居製麺 たぬき物産				
3	1950	250	250	250	250
4	1955	292	308	285	313
5	1960	324	339	315	317
6	1965	351	367	322	352
7	1970	413	394	363	420
8	1975	482	421	367	518
9	1980	595	430	432	528

このような状態で入力された場合、データが入力されている列の中のセルをアクティブにしてから[データ]-[区切り位置]を使うことによって、「区切り位置ウィザード」を起動することができる。



3. おまけの情報

日本の官庁が作成している統計情報については、総務省統計局で公開状況などを整理して公表している。<http://www.stat.go.jp/data/guide/index.htm> を参照

練習問題 4

福島県白河市の事業所数・従業者数について以下の点をわかりやすく図あるいは表に示しなさい

- (1) 民営事業所数（産業計）の時系列変化
- (2) 公営と民営事業のシェア時系列変化（産業計）を事業所数と従業者数のそれぞれについて
- (3) 白河市の産業構造の変化