

## ピボットテーブル(Pivot Table)

- データを表の形で集計、分析するための強力なツール
- fact finding から営業日報作りにまで使える汎用性
- マイクロデータを扱う場合、Excel のなかでも特に「使える」機能

### 0. ピボットテーブルで扱うデータの構造

#### 0.1 データの並びとフィールド名

ピボットテーブルウィザードは、1 系統のデータはワークシートの列方向に（つまり、A 列は品目、B 列は売上げ個数...といった感じ）並んでいることを大前提としている。これは Excel の他の機能と全く同じである。

ピボットテーブルに特有の制限は、データ範囲の一番上の行（先頭行）には、データ系列の名前を表す文字列を入れる必要があることである。この名前を「フィールド名」と呼ぶ。ピボットテーブルを作成するために使うウィザードは、フィールド名でデータ系列を管理するため、

- 名前が入っていない場合はそもそもウィザードに処理を断られる
- 複数の系列に同じ名前が付いているのも当然ダメ
- 実際のウィザード画面は図 4 のようになるため、長すぎる名前は末尾が切れてしまうなどの制限もあり、名前をうまくつける<sup>1</sup>のは作業効率を上げるためにはかなり重要。

フィールド名 1	フィールド名 2	フィールド名 3	フィールド名 4
データ			
データ			
データ			

### 1. ピボットテーブルウィザードの使い方

#### 1.1 ピボットテーブルウィザードの起動

ピボットテーブルを作成する場合は、[データ(D)]-[ピボットテーブルレポート(P)]から、ピボットテーブルウィザードを起動する（図 1）。Excel2000 以降の場合は、あまり使っていないサブメニューが表示されなくなっているので必要に応じてメニュー展開ボタン（下向きの三角が 2 つ重なったようなアイコン）を押すこと。

<sup>1</sup> 名前を上手く付けることが重要なのは別にこれに限った話ではない。プログラミングにおける関数名もそうだし、新しい概念にインパクトのある名前をつけることはとても大事。

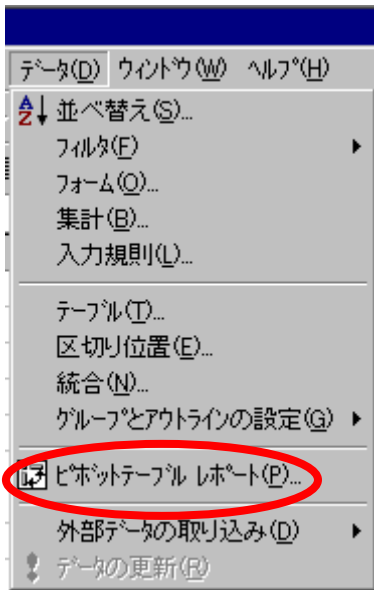


図 1 ピボットテーブルレポートの起動

### 1.2 データソースの指定

ウィザードを起動すると、最初にデータをどこから持ってくるかを指定する画面がでてくる(図 2)。外部データベースから持ってくるというのも実務的には重要なことが多いが(例えば売上げ管理データベースにアクセスできる権限があれば、直接必要なデータを呼び出して Excel で分析できる)、データベース自体に関する知識が必要なのでこの授業では扱わない。初期のデータソースは「Excel のリスト・データベース」となっているので、このままそれを使う。

### 1.3 データ範囲の指定

Excel のリスト・データベースを選択した場合は、データが入っている範囲を図 3 の画面で指定する。例では列全体を指定しているが、このような指定をするとデータが入っている範囲を Excel が勝手に判断して使う。

### 1.4 集計用フィールドの指定

データとして選択した範囲に入っているフィールドのうち、どのフィールドを集計の表頭、表側、データとして使うかを指定する画面(図 4)。要するにクロス表の内容を指定する一番重要な画面である。操作方法は、ドラッグ&ドロップが基本。フィールド名をドラッグして、目的の部分にドロップすれば図 5 のような形で指定できる。ここでフィールド名の表示が長すぎると切れてしまっていることに注意。例えば内申平均点は「内申平均」としか表示されていない。フィールド名を付けるときは見える範囲で区別できるような名前をつけることが重要なことがわかる。

### 1.5 集計方法の指定

図 5 を見るとわかるように、デフォルトの集計方法は「データの個数」である。場合によ

ってはこのままで構わないが、今回の集計では国語の平均点が出したいので個数がでもあまりうれしくない。この場合は、集計フィールド指定のデータをダブルクリックして変更することができる(図 6)。図ではスクロールアウトして見えないが、指定可能な集計方法は(1)合計(2)データの個数(3)平均(4)最大値(5)最小値(6)積(7)数値の個数(8)標本標準偏差(9)標準偏差(10)標本分散(11)分散と多様である。適切な物を使うようにしたい。

このようにウィザードに従って順次指定していくと、図 7 のようなクロス集計表が完成する。図 7 は出身校の種類(A, B, C)と内申平均点と、国語の試験の成績の平均を集計したものである。

#### 1.6 集計内容の変更

発見的なデータ処理を行う場合は、いろいろなデータ間の相関を見たい。そのためにはデータの組み合わせを変えつつクロス表を山ほど作る必要がある。このとき、ウィザードを最初から動かすのは冗長であるしデータ範囲の指定を間違えたりすると面倒であるから、図 4 にすぐ戻れると便利。

いつものようにピボットテーブルで作成したクロス表を選択してから右ボタンクリックをすると、図 8 のようなメニューが表示される。このなかのウィザードを選択すると、フィールド名の選択に戻れる。

#### 1.7 複数の集計を同時に行う

同じ表頭と表側で複数のデータを同時に集計することができる。図 9 は例として複数科目の平均点を同時に計算しているが、実際のデータ解析のときには平均や標準偏差のような特性値とデータの個数を同時に計算するのが一番役に立つ。個数が少ない場合は特性値自体が当てにならないが、特性値だけ計算した場合直観的にそのことが分かりづらくなる。

#### 1.8 多重クロス集計を行う

表頭あるいは表側で複数のカテゴリを組み合わせる集計を行う場合は、複数のフィールド名を表頭あるいは表側部分にドロップすればよい(図 11)。このような操作を行うと多重クロス表が作成できる(図 12)。ただし、あまりやりすぎると見づらい表ができるので注意。

#### 練習問題 1

pivot.xls への「成績」データはある高校での成績データである。内申平均点は、この高校の学区内にある 3 中学校(A, B, C)での成績(相対評価)を平均したものとなっている。科目毎の成績は高校で実施した試験の結果(絶対評価)を示している。

1-1 出身中学ごとに、内申平均点の分布をヒストグラムに図示しなさい

1-2 5 科目の平均得点の分布を中学毎に分けて同一のヒストグラムに図示しなさい

1-3 高校で実施した試験結果から考えて、学区内の 3 つの中学の学力水準には差があると考えられるか。また差があると考えられる場合にはどのように差があるのかを示しなさい。

#### 練習問題 2

povot.xls の「就業行動」は高齢者の就業パターンに関する調査データである。このデータ

から高齢者の就業にはどのような要素が効いていそうか考察しなさい。

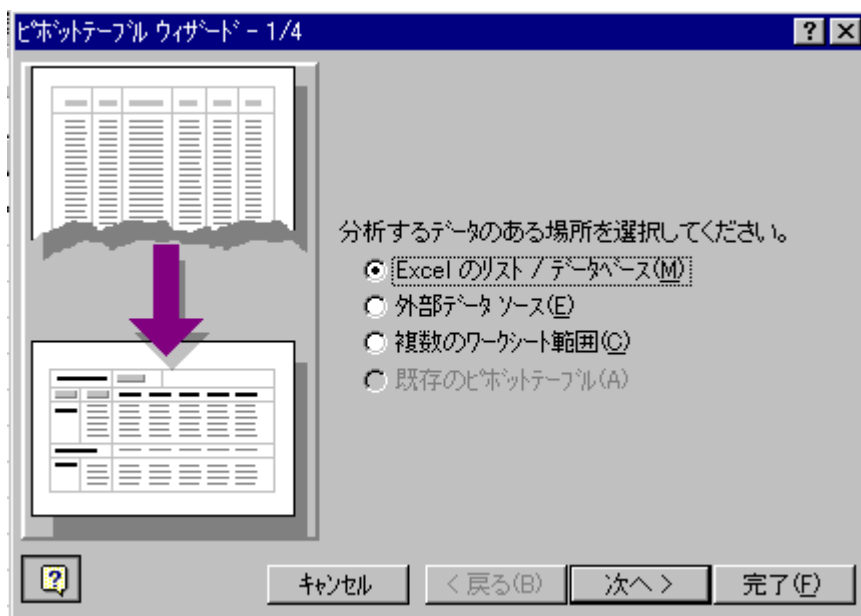


図 2 ピボットテーブルのデータソースの指定

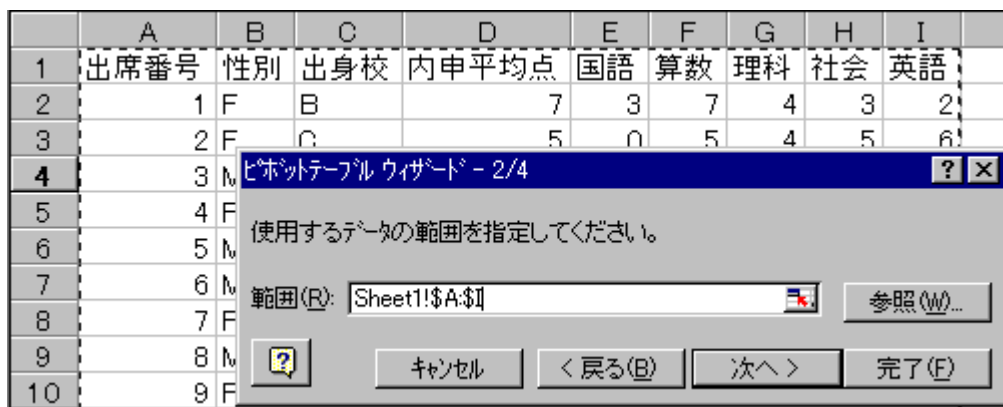


図 3 Excel のリストからデータを指定する場合

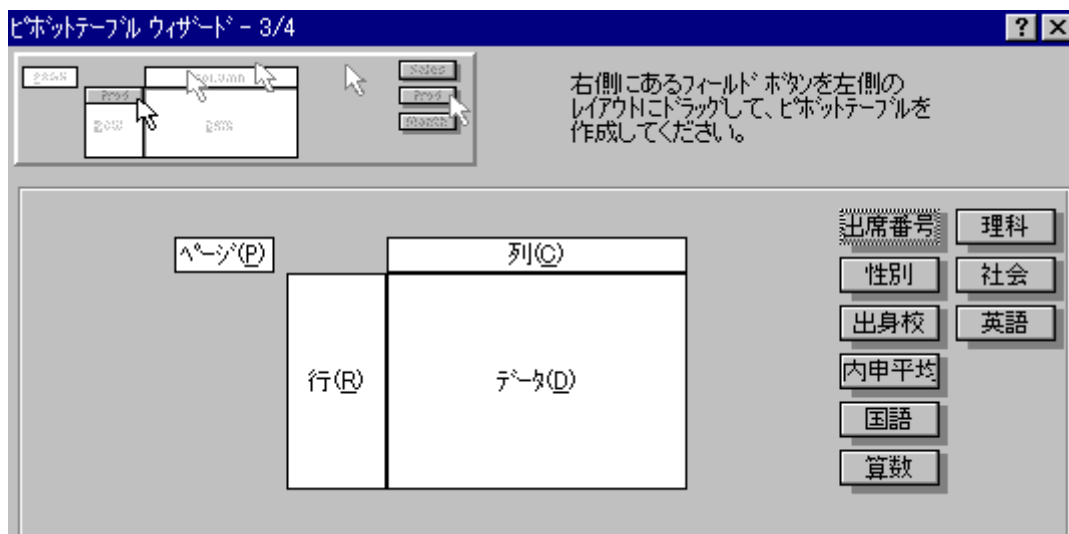


図 4 集計対象フィールドの指定

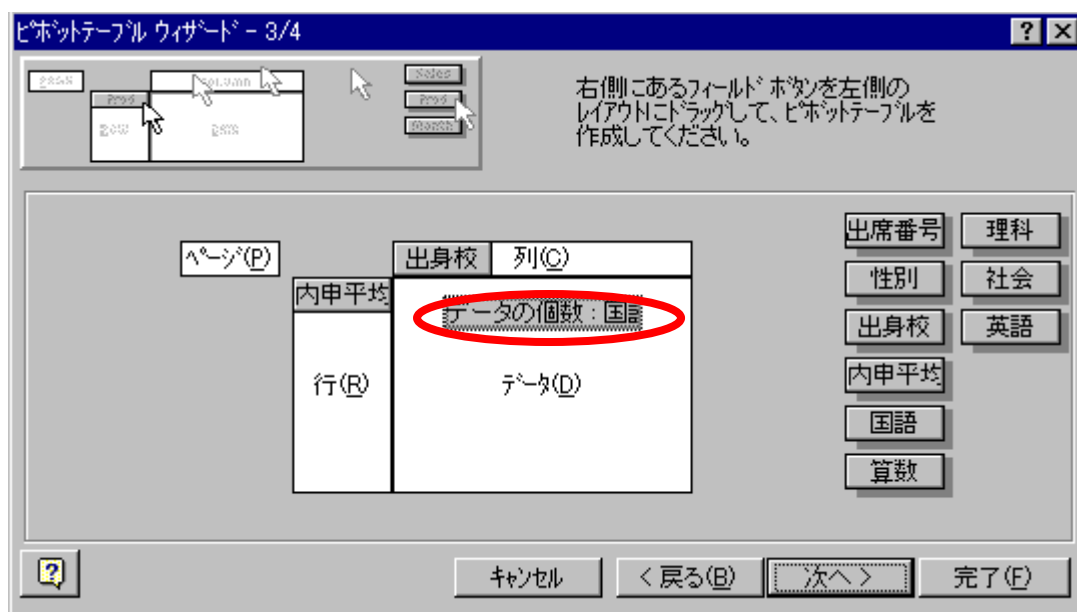


図 5 集計方法を変更するには、データフィールドをダブルクリック



図 6 集計方法の指定

平均: 国語	出身校				
内申平均点	A	B	C	(空白)	総計
1	7			5	6.333333333
2	4.333333333	2.666666667		3	3.133333333
3	6.833333333		2	2.25	3.523809524
4	4.125	3.833333333	3.076923077		3.555555556
5	4		3.5	3	3.5
6	4.333333333	3.727272727	2.923076923		3.5
7	4.8	3.428571429		3	3.75
8	3		4.8	3.5	3.857142857
9	4	4.333333333	1.333333333		3.222222222
10	6		2	5	4.285714286
(空白)					
総計	4.596491228		3.5	2.987654321	3.605

図 7 作成されたクロス集計表の例



図 8 クロス集計表を右ボタンクリックで表示されるサブメニュー



図 9 集計項目に複数のフィールドを指定できる

		出身校				
内申平均点	データ	A	B	C	(空白)	統計
1	平均: 国語	7			5	6.333333333
	平均: 英語	6.5			5	6
2	平均: 国語	4.333333333	2.666666667		3	3.133333333
	平均: 英語	5.666666667	3.833333333	1.333333333		3.2
3	平均: 国語	6.833333333		2	2.25	3.523809524
	平均: 英語	6.166666667	3.666666667		3.25	4.142857143
4	平均: 国語	4.125	3.833333333	3.076923077		3.555555556
	平均: 英語	5.75	4.333333333	2.153846154		3.703703704
5	平均: 国語	4		3.5	3	3.5
	平均: 英語	4.666666667	4.083333333	2.533333333		3.738095238
6	平均: 国語	4.333333333	3.727272727	2.923076923		3.5
	平均: 英語	4.833333333	4.080909091	2.923076923		3.733333333
7	平均: 国語	4.8	3.428571429		3	3.75
	平均: 英語	4.3	3.785714286		2.5	3.625
8	平均: 国語	3		4.8	3.5	3.857142857
	平均: 英語	4		5.6	3	4.142857143
9	平均: 国語	4	4.333333333	1.333333333		3.222222222
	平均: 英語	5		3	1.333333333	3.111111111
10	平均: 国語	6		2	5	4.285714286
	平均: 英語	9		5.5	2.5	4.285714286
(空白)	平均: 国語					
	平均: 英語					
全体の	平均: 国語	4.596491228		3.5	2.987654321	3.606
全体の	平均: 英語	5.105263158	4.112903226	2.567901235		3.77

図 10 複数の集計フィールドを含んだクロス表



図 11 表頭項目に2つのフィールドを指定する

平均: 国語	出身校	性別				
内申平均点	A		A 計	B		B 計
	F	M		F	M	
1			7	7		
2		3	5	4.333333333	2.333333333	3
3	5.666666667		8	6.833333333	1	2.5
4	5.5	2.75	4.125	4.333333333	3.333333333	3.8333
5	4.625	3.285714286	4	4.2	3	
6	4.5	4	4.333333333	3.714285714	3.75	3.7272
7	4.75	5	4.8	3.571428571	3.285714286	3.4285
8	1	7	3	5	4	
9	8	2	4	6	3.5	4.3333
10	6		6	2		
(空白)						
総計	4.71875	4.44	4.596491228	3.727272727	3.24137931	

図 12 表頭項目が複合項目になった多重クロス表