

ピボットテーブル(Pivot Table)の使い方

- 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。

	A	B	C	D	E
1	生徒番号	性別	出身校	国語	算数
2	1		A校	65	10
3	2	M	C校	60	16
4	3	F	A校	57	44
5	4	F	A校	72	52
6	5	F	B校	25	50
7	6	F	B校	73	83
8	7	F	A校	53	4

「生徒番号」データの並び

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
- 長すぎる名前は画面表示上末尾が切れてしまうことがある

などの制限もあり、名前をうまくつける¹のは作業効率を上げるためにはかなり重要。

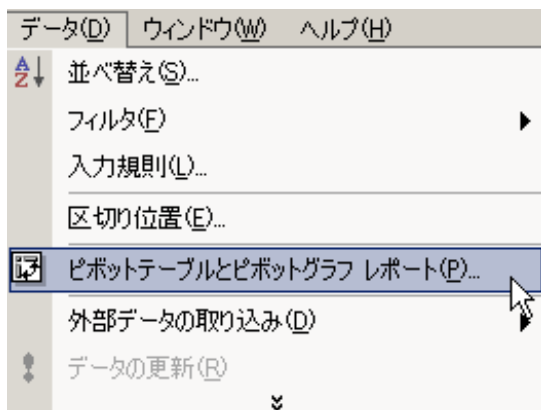
フィールド名 1	フィールド名 2	フィールド名 3	フィールド名 4
データ			
データ			
データ			

¹ 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。

例題 1. score.xls のデータから、表頭出身校、表側に性別を配置して英語の平均点を計算したクロス集計表を作成してみる。クロス集計表をつくるためには、ピボットテーブルウィザードを利用する。

ピボットテーブルウィザードの起動

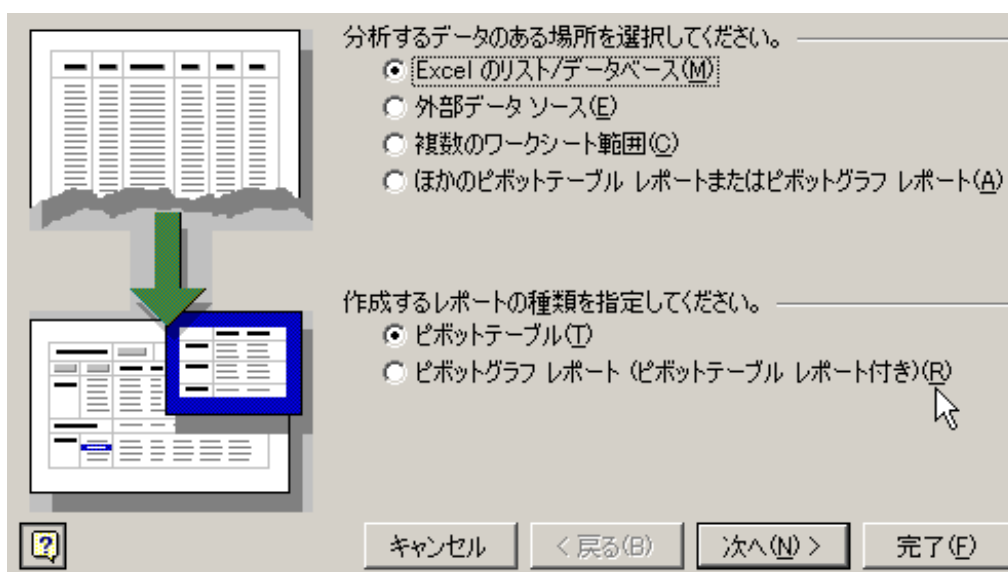
ピボットテーブルを作成する場合は、[データ(D)]-[ピボットテーブルとピボットグラフレポート(P)]から、ピボットテーブルウィザードを起動する。



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。

データソースの指定

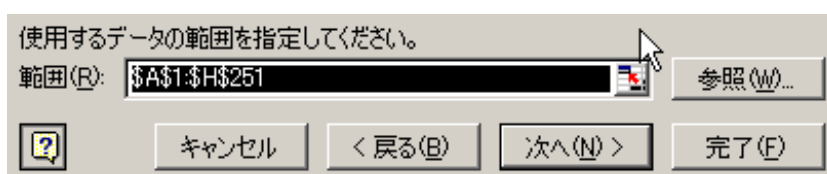
ウィザードを起動すると、最初にデータをどこから持ってくるかを指定する画面がでてくる。



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。

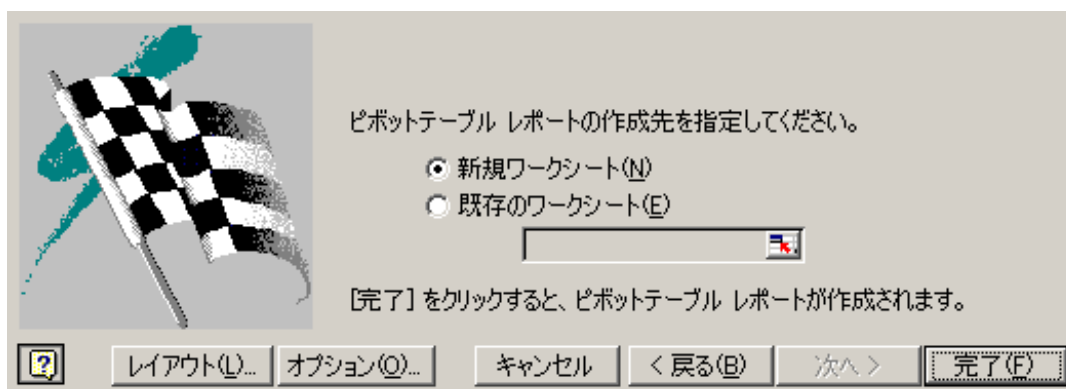
データ範囲の指定

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範囲指定で指定する。指定方法はグラフなどのデータ指定と同じ。ただし、フィールド名が先頭行になるように指定すること。



ピボットテーブル作成位置の指定

ピボットテーブルを新規に作成したワークシート上に作るか、既存のワークシート上に作るかを指定する。既存のワークシート上に作る場合は、左上の位置をセル番地で指定する。既存のワークシート上に作る場合は、右方向と下方向が空いている場所を指定すること。時々フィールドの指定違いなどで妙に大きな集計表ができてしまうことがあるが、右と下が空いていればデータを壊す心配がない。



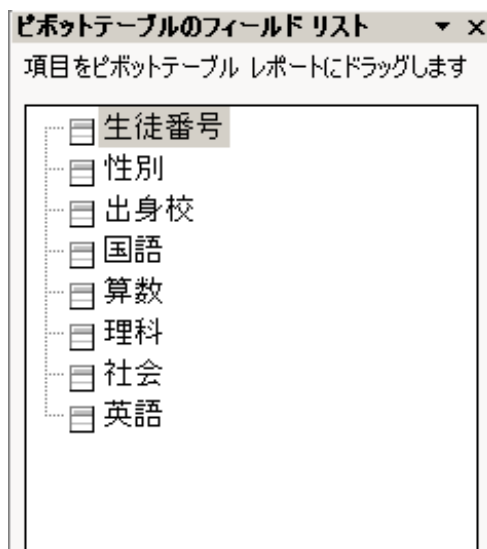
集計用フィールドの指定

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。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指定できる。「列のフィールド」とは、表頭に使うフィールド名を

ドロップする部分、「行のフォールド」とは表側に使うフィールド名をドロップする部分、そして、「データアイテム」とは、集計する時に計算に用いるフィールド名をドロップする部分である。「ページのフィールド」は、フィールドの内容によって集計表を複数ページに分割したい場合に、分割に使うフィールド名をドラッグ&ドロップする。

	A	B	C	D	E	F	G
1	ここにページのフィールドをドラッグします						
2							
3	ここに列のフィールドをドラッグします						
4	ここに行のフィールドをドラッグします	ここにデータ アイテムをドラッグします					
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							

ドロップすべきフィールド名のリストは、下図のようなリストとして提示されている。ここからドラッグ&ドロップすればよい。



今回は、表頭が「出身校」、表側が「性別」、集計するのは「国語」の平均点 であるから、それぞれ「出身校」を「列のフィールド」、「性別」を「行のフィールド」、「国語」を「データアイテム」にドロップする。すると

	A	B	C	D	E
1	ここにページのフィールドをドラッグします				
2					
3	合計 / 国語	出身校			
4	性別	A校	B校	C校	総計
5	F	1846	3284	2617	7747
6	M	1412	2111	2218	5741
7	総計	3258	5395	4835	13488
8					

合計を計算
している

のような表ができる。確かに表頭と表側は正しいようだが、中のデータはどうみても平均点には見えない。よくよく見ると、左肩のセルに「合計 / 国語」と書いてある。つまり、平均ではなく合計点を計算しているのである。

集計方法の変更

この場合欲しいのは平均点であるから、合計点のままでは困る。このように集計方法が希望するものと Excel の Wizard が判断したものと異なる場合²、上図赤丸で囲んだ集計方法の書いてあるセルをダブルクリックすることによってメニューが開く。

フィールド名: 国語
 名前(M): 合計 / 国語
 集計の方法(S):

- 合計
- データの個数
- 平均
- 最大値
- 最小値
- 積
- 数値の個数

OK
 キャンセル
 表示しない(H)
 表示形式(N)...
 オプション(O) >>

可能な集計方法は合計、データの個数、平均、最大値、最小値、積、数値の個数、標本標準偏差、標準偏差、標本分散、分散 などである。ここでは平均を計算したいので「平均」を選んで「OK」を押すと、

	A	B	C	D	E
1					
2					
3	平均 / 国語	出身校			
4	性別	A校	B校	C校	総計
5	F	52.74285714	56.62068966	59.47727273	56.54744526
6	M	52.2962963	47.97727273	52.80952381	50.80530973
7	総計	52.5483871	52.89215686	56.22093023	53.952
8					

今度は正しく平均 / 国語 になっており、データもいかにも平均値らしいデータになって

²既に皆さん予想がつくように、このケースでも Excel の Wizard は結構間抜けな判断をする。あまり信じないように。

いる。

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、フィールド名の指定にすぐ戻れると便利。

ピボットテーブルで作成したクロス表をクリックすると、フィールド名のリストが再表示される。集計に不要になったフィールドは、集計表からドラッグしてフィールド名のリストにドロップすれば取り除くことができる。データアイテムは左肩のセルをドラッグしてフィールド名リストにドロップすることで取り除ける。


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。たとえば平均や標準偏差のような特性値だけ計算した場合、データの個数が少ないと実際問題としてあまり意味がない。しかし、表にデータの個数が入っていないとそのことが直観的に分かりづらく、1000個データがある場合の平均と1件しかないデータの「平均」を同様に扱ってしまう危険がある。

このようなケースでは、「データアイテム」に2回同じデータをドロップすることによって平均と件数を同じ表に入れることができる。

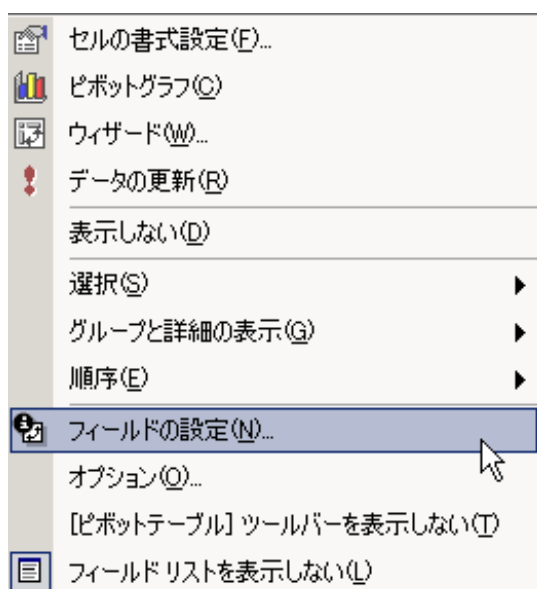
は例として複数科目の平均点を同時に計算しているが、実際のデータ解析のときには平均や標準偏差のような特性値とデータの個数を同時に計算するのが一番役に立つ。個数が少ない場合は特性値自体が当てにならないが、特性値だけ計算した場合直観的にそのことが分かりづらくなる。

	A	B	C	D	E
1					
2					
3	平均 / 国語	出身校			
4	性別	A校	B校	C校	総計
5	F	52.74285714	56.62068966	59.47727273	56.54744526
6	M	52.2962963	47.97727273	52.80952381	50.80530973
7	総計	52.5483871	52.89215686	56.22093023	53.952
8					

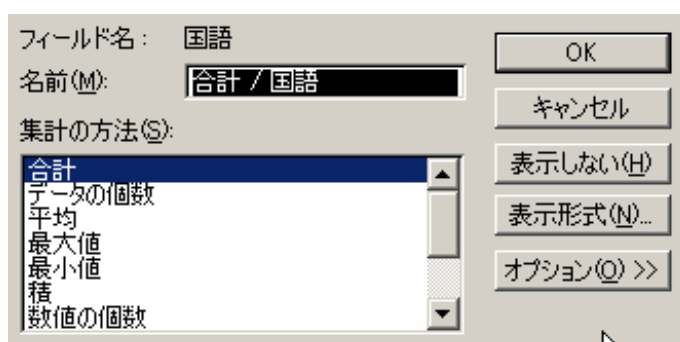
のような集計表があった場合、国語のデータ件数を表に追加するためには、フィールド名リストから「国語」をもう一度「データアイテム」部分、すなわち上図の  の範囲にドロップすればよい。上の表に国語をもう一度ドロップすると、

	A	B	C	D	E	F
1						
2						
3			出身校			
4	性別	データ	A校	B校	C校	総計
5	F	平均 / 国語	52.74285714	56.62068966	59.47727273	56.54744526
6		合計 / 国語	1846	3284	2617	7747
7	M	平均 / 国語	52.2962963	47.97727273	52.80952381	50.80530973
8		合計 / 国語	1412	2111	2218	5741
9	全体の 平均 / 国語		52.5483871	52.89215686	56.22093023	53.952
10	全体の 合計 / 国語		3258	5395	4835	13488
11						

のような表ができる。このままだと2回目にドロップしたデータが「合計 / 国語」になっているため集計方法を変更する必要がある。この場合は、「合計 / 国語」のセル（上図赤囲み）を右ボタンクリックして出てくるダイアログ



から「フィールドの設定」を選ぶことによって、



の設定ダイアログが再度呼び出せるので、これを用いて適宜集計方法を変更すればよい。ここでは「データの個数」を選ぶことによって、

3			出身校			
4	性別	データ	A校	B校	C校	総計
5	F	平均 / 国語	52.74285714	56.62068966	59.47727273	56.54744526
6		データの個数 / 国語	35	58	44	137
7	M	平均 / 国語	52.2962963	47.97727273	52.80952381	50.80530973
8		データの個数 / 国語	27	44	42	113
9	全体の平均 / 国語		52.5483871	52.89215686	56.22093023	53.952
10	全体のデータの個数 / 国語		62	102	86	250

のように平均と個数の両方が入ったクロス集計表を作成することができる。

練習問題 1.

例題と同じ score.xls を用いて、まず国語、算数、理科、社会、英語の合計点を計算し、例題と同じように出身校および性別ごとの平均点とデータの個数が入ったクロス集計表を作成しなさい。

練習問題 2.

Excel のピボットテーブルレポートは、連続データを区間に区切ってカテゴリ化する機能を持っていないため、点数のようなデータを直接ピボットテーブルの表頭や表側のフィールドに指定すると大変なことになる。このことを表頭に英語の得点、表側に国語の得点を入れて人数を求めるピボットテーブルを作って確認しなさい。

練習問題 3

連続データを一定区間に区切ってカテゴリ化する方法はいくつか考えられる。たとえば IF 関数を用いてカテゴリ化する、除算と INT 関数を組み合わせて階段状のデータに変換してカテゴリ化するなどの方法がすぐ思いつく。

練習問題 2 では素点をそのまま使ったが、10 点階級のカテゴリデータに変換して英語と国語の得点階級ごとの人数を求めなさい。

Tips. 元データの確認

Excel で作成したピボットテーブルは元のデータとの関係を知っているため、集計表の特定のセルに入っているデータが実際にはどんなデータであるかを表示することが可能である。たとえば、例題 1 の最終的な結果表で A 校出身の男性のデータを見たければ、A 校男性のデータ（平均でもデータの個数でもよい）をダブルクリックすれば新しいワークシートが挿入され、そこに元のデータが全て表示される。この機能は、特異な外れデータが出たときに、データの中身をすぐ確認できて便利。

Tips. 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て集計を行う場合は、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。このような操作を行うと下図のような多重クロス表が作成できる。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。単にデータを分割したいだけで、別の表になっていてもかまわないなら、「ページのフィールド」を使って複数の表に分割した方が見通しはよくなる。

2									
3	データの個数 / 生徒番号	性別 ▼	出身校 ▼						
4		F			F 合計	M		M 合計	(空白)
5	合計点階級 ▼	A校	B校	C校		A校	B校	C校	(空白)
6	5						1		1
7	7						1		1
8	8					1			1
9	9							1	1
10	10						1	1	2
11	11					1		1	2
12	12	1			1				
13	13	1	1		2	4	1		1
14	14		1		2	3	1	1	2
15	15	3			1	4	1	2	3
16	16	1				1	2	2	4
17	17		3		4	7	1	1	2
18	18	1	4		1	6	1	1	2
19	19	1			1	2	1	2	1
20	20		4		4	8	2		4
21	21	1			2	3		4	1
22	22	2	3		3	8	2	3	4
									9

練習問題 4.

pivot.xls 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。

4-1 出身中学ごとに、内申平均点の分布を横軸を内心平均点、縦軸を人数としたグラフに図示しなさい（ヒント ピボットテーブルで作成した表を加工する場合は、「形式を選択して貼り付け」機能で「値」だけ複写すると扱いやすい）。

4-2 5 科目の平均得点の分布を中学毎に分けて同一のグラフに図示しなさい。ただし、平均得点分布は 1 点階級で描くこと。

4-3 高校で実施した試験結果が正しく生徒の学力を評価できていると仮定した場合、学区内の 3 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。