

タイトル	非正規形データの統計解析における情報システムの活用について
著者	穴沢, 務
引用	北海学園大学学園論集, 120: 39-71
発行日	2004-06-25

非正規形データの統計解析における 情報システムの活用について

穴 沢 務¹

1. はじめに

本稿では、本来ならば関係データベース管理システム (RDBMS) による処理が適しているデータを、表計算ソフトで、しかも RDBMS から見て不適切な形式で格納し統計処理する場合に生じる問題点と、その克服方法について検討する。

今日、統計処理を目的とした情報システムの種類は、扱えるデータの大小、手法の難易度、利用の手軽さなどの面から見て多岐に及んでいる。中でも、入手のしやすさや使いやすさから多方面で普及している情報システムとして、表計算ソフトでは Microsoft Excel (以下 Excel)、RDBMS では Microsoft Access (以下 Access) が挙げられよう。そこで本稿では、統計処理においてそれら 2 つ (とりわけ前者) を利用するという前提で議論を進めたい。

一般に、Excel は任意のセルを用いた柔軟な表計算が可能なこと、数値計算や統計処理のための関数やツールが豊富に揃っていることなどから、Access に比べて高度な統計処理が可能である。一方 Access は、いくつかの表に分割して格納した方が良いデータ (例えば内訳を含む伝票の集合からなるデータ) の扱いに優れている。したがって、2次元のセル空間にフラットに格納されているデータだけを処理する場合は Excel を用い、そうでない場合は Access で複数の表を適切な形式に結合してから統計処理を行う (さらに高度な処理が必要なときは、結合した表を Excel ファイルとしてエクスポートする) のが常套手段となっている。

最初に述べた「RDBMS による処理が適しているデータ」とは、格納時にはいくつかの表に分割し、必要に応じて表を結合して利用するようなデータのことである。しかし、実際の統計処理の現場では、そのようなデータを Excel のワークシートに 1つの表として無理やり格納しているケースをよく見かける。そのような状況に対する最も望ましい改善策は、RDBMS を併用した統計解析を推奨することである。しかし現実には、以下のような理由からそのような切り替えができないことがある。

1 本稿で紹介する技法には、本学の福永厚教授との共同作業の中で知りえたものが多く含まれる。福永教授にはこの紙面を借りて厚く御礼を申し上げたい。

- RDBMS は Access のような GUI 対応ソフトであっても、使いこなせるようになるまでにか
なりのトレーニングを要する。
- 在任期間が短い部署や、情報システムに不慣れな人（高齢者など）が多い部署では、そのよ
うなトレーニングが事実上不可能であり、Excel のように多くの人が使い慣れたソフトを使
わざるを得ない。
- 過去にそのような形式で蓄えられたデータを再利用することが多い。

このようなネガティブな理由以外にも、次のようなポジティブな理由からあえて RDBMS から
見て不適切な形式を温存する場合もある。

- すべてのデータが1つの表にまとまっていると全体像を見渡す場合に便利である。

そこで本稿では、そのようなやむを得ない状況下での Excel による統計解析の方法について考
察する。方法論を展開する際の基本方針として、上記のような「情報システム弱者」のスキルアッ
プを図るという観点から、多種多様な技法を網羅することより、少数の汎用的な技法の組み合わ
せによる手法の模索に重点を置く。とはいえ、完全な初心者を導くことは紙面の制約から見て不
可能なので、すでに次のようなことが実行できる読者を想定する。

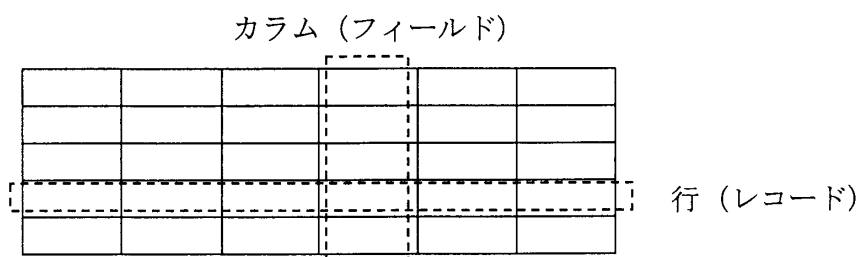
1. Windows の基本操作（マウスによる窓の開閉・伸縮・移動、応用ソフトの起動など）ができ
ること。
2. ファイル管理（ファイルの複写・移動・削除、フォルダの作成、ファイルの検索）ができ
ること。
3. Excel の基本操作（データ・計算式・関数の入力、セル内容の複写、セル書式の変更）でき
ること。

本稿では、本来ならば RDBMS での処理が望ましいデータの統計処理について考えるので、ま
ず第2章で関係データベースの概念について簡単に触れ、RDBMS の観点から望ましくないデー
タの格納形式（非正規形）について述べる。第3章では、本稿が想定する統計処理について述べ、
それらを実行するための Excel の技法を第4章で紹介する。第5章では、第4章で紹介する技法
の中のいくつかを非正規形データに適用して（第3章で述べる）統計処理を実行するための具体
的な方法とその限界について検討する。

2. 関係データベースについて

関係データベースとは、1970年に E. F. Codd が提案したデータモデルであり、その最も基本
的な要件は、すべてのデータを図2.1のような2次元の**テーブル（表）**に格納することである。

ここで縦の各列を**カラム**（または**フィールド**）、横の各列を行（または**レコード**）という。デー
タベースに格納する物件は、通常いくつかの属性（項目）によって特徴付けられるが、1物件の
情報はすべて1つの行に格納し、ある属性に対する各物件の値は同じカラムに格納することを約
束する。例えば、2人の人物について4種類の属性（氏名、性別、身長、体重）に対する値がわ



(図 2.1) テーブル

かっているとき、それらは図 2.2 のように格納される。

氏名	性別	身長	体重
鈴木	女	158	45
山田	男	175	67

(図 2.2) データの格納例

図 2.2 の例ではテーブル内の各セルに対しそれぞれ 1 つの値を格納するような形式をとっているが、1 つのセルにさらに表形式のデータを内包するような場合も存在する。図 2.3 のような 2 枚の売上傳票をデータベース化することを考えよう。

<p>NO 0001 2004/5/7</p> <p style="text-align: center;">豊平商事 様</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th colspan="4">内訳</th></tr> <tr><th>品名</th><th>単価</th><th>数量</th><th>小計</th></tr> <tr><td>シャープペン</td><td>100</td><td>5</td><td>500</td></tr> <tr><td>消しゴム</td><td>60</td><td>5</td><td>300</td></tr> <tr><td>修正テープ</td><td>400</td><td>2</td><td>800</td></tr> <tr><td colspan="3">合計</td><td>1600</td></tr> <tr><td colspan="3">消費税</td><td>80</td></tr> <tr><td colspan="3">税込み</td><td>1680</td></tr> </table>	内訳				品名	単価	数量	小計	シャープペン	100	5	500	消しゴム	60	5	300	修正テープ	400	2	800	合計			1600	消費税			80	税込み			1680	<p>NO 0002 2004/5/7</p> <p style="text-align: center;">旭町工業 様</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th colspan="4">内訳</th></tr> <tr><th>品名</th><th>単価</th><th>数量</th><th>小計</th></tr> <tr><td>ボールペン黒</td><td>100</td><td>10</td><td>1000</td></tr> <tr><td>セロテープ大</td><td>220</td><td>5</td><td>1100</td></tr> <tr><td colspan="3">合計</td><td>2100</td></tr> <tr><td colspan="3">消費税</td><td>105</td></tr> <tr><td colspan="3">税込み</td><td>2205</td></tr> </table>	内訳				品名	単価	数量	小計	ボールペン黒	100	10	1000	セロテープ大	220	5	1100	合計			2100	消費税			105	税込み			2205
内訳																																																													
品名	単価	数量	小計																																																										
シャープペン	100	5	500																																																										
消しゴム	60	5	300																																																										
修正テープ	400	2	800																																																										
合計			1600																																																										
消費税			80																																																										
税込み			1680																																																										
内訳																																																													
品名	単価	数量	小計																																																										
ボールペン黒	100	10	1000																																																										
セロテープ大	220	5	1100																																																										
合計			2100																																																										
消費税			105																																																										
税込み			2205																																																										

(図 2.3) 伝票形式のデータ

この場合、1 枚の伝票を 1 物件と見なせば、これらのデータを格納するには 3 つのカラム「No」「日付」「顧客名」「内訳」が必要となるが、これでは「内訳」カラムにさらに表形式のデータを含むことになり、その後の分析が困難になることは想像に難くない。そこで、1 物件は 1 行に、しかし 1 セルには 1 つの値という原則を貫く方法として、図 2.4 のような格納形式が考えられる。(但し、後から計算可能な「小計」「合計」「消費税」「税込み」はテーブルから除く。)

NO	日付	顧客名	品名 1	単価 1	数量 1	品名 2	単価 2	数量 2	品名 3	単価 3	数量 3
0001	2004/5/7	豊平商事	シャープペン	100	5	消しゴム	60	5	修正テープ	400	2
0002	2004/5/7	旭町工業	ボールペン黒	100	10	セロテープ大	220	5			

(図 2.4) 非正規形

この形式の特徴は、カラムとして「内訳」ではなくその細目である「品名」「単価」「数量」を設け、さらに内訳の最大件数だけそれらを横方向に並べている点である。これら「品名」「単価」「数量」のように、横方向に繰り返し現れるカラムを**繰り返しフィールド**といい、繰り返しフィールドを用いたデータの格納形式を**非正規形**という。なお、繰り返しフィールド以外のカラム(図2.4では「NO」「日付」「顧客名」)をここでは**非繰り返しフィールド**と呼ぶことにしよう。本稿では、図2.4のような非正規形データに対する統計処理について考察する。しかし、非正規形データは内訳の件数によってカラム数が変わるので、Excelなどで定型的な処理を行うことが難しいなどの欠点がある。また、第5章で示すように非正規形データに対する統計処理にはさまざまな困難が伴う。

一方、図2.4のような繰り返しフィールドを用いずに、データを1つのテーブルに格納する方法もある。図2.5では、内訳の中の「品名」「単価」「数量」にそれぞれ1カラムずつ割り当て、(伝票1枚ではなく)内訳1件を1物件と見なしてデータを格納している。このような繰り返しフィールドを設けないデータの格納方法を**第1正規形**という。この形式は、Excelなどで特に内訳の中のデータについて統計処理する場合に優れており、第5章でも非正規形データとの比較として第1正規形データの処理例を一部紹介する。一方で、第1正規形では各行がどの伝票のデータであるかを識別するために、内訳以外の項目である「NO」「日付」「顧客名」のカラムに重複値を格納している。この点は、入力ミスを誘発したりディスク領域を無駄に消費するという点で好ましいことではない。

NO	日付	顧客名	品名	単価	数量
0001	2004/5/7	豊平商事	シャープペン	100	5
0001	2004/5/7	豊平商事	消しゴム	60	5
0001	2004/5/7	豊平商事	修正テープ	400	2
0002	2004/5/7	旭町工業	ボールペン黒	100	10
0002	2004/5/7	旭町工業	セロテープ大	220	5

(図2.5) 第1正規形

第1正規形の欠点を補う格納形式として、図2.6のようにデータを2つのテーブルに分割する方法がある。この形式では、内訳の情報を図2.6右のテーブルにまとめ、各内訳がどの伝票に属するかは「NO」カラムで照合できるようになっている。この格納形式を**第2正規形**という。さらに、図2.6左の「顧客名」カラムと図2.6右の「品名」カラムには重複値が現れる可能性が高いことから、図2.7のような更なる細分化を図る場合がある。図2.7のレベルまでデータを細分化して格納する形式を**第3正規形²**という。RDBMSでは、データを第2もしくは第3正規形で格納

2 本稿での第2、第3正規形の説明は文献[3]に従った。しかし、正規形の定義については実用を重視した文献([3]など)と学術色の強い文献([1]、[2]など)の間で若干の差異が見られる。このような用語の定義に関する整理は、本稿の目的を越えるので別の機会に行いたい。

NO	日付	顧客名	NO	品名	単価	数量
0001	2004/5/7	豊平商事	0001	シャープペン	100	5
0002	2004/5/7	旭町工業	0001	消しゴム	60	5
			0001	修正テープ	400	2
			0002	ボールペン黒	100	10
			0002	セロテープ大	220	5

(図 2.6) 第2正規形

NO	日付	顧客コード	顧客コード	顧客名	NO	品名コード	単価	数量	品名コード	品名
0001	2004/5/7	T 1	T 1	豊平商事	0001	S 1	100	5	S 1	シャープペン
0002	2004/5/7	A 1	A 1	旭町工業	0001	K 1	60	5	K 1	消しゴム
					0001	S 2	400	2	S 2	修正テープ
					0002	B 1	100	10	B 1	ボールペン黒
					0002	S 3	220	5	S 3	セロテープ大

(図 2.7) 第3正規形

し、必要に応じてテーブルを図 2.5 の形式に結合して分析を行うことが多い。しかし Excel では、VLOOKUP 関数を用いた同様なテーブル結合は可能なものの、RDBMS に比べるとその手間は煩雑なので、分割した格納したデータを統計処理する際には多少の困難が伴う。

このように、データの処理効率と格納効率を考えながらデータを細分化するための方法論を**正規化理論**という。

3. 集計の種類

本稿では、統計処理専用ソフトではない RDBMS でも実行可能であるような基本的な統計処理として集計のみを考察する。そこで、この章において集計の種類について整理をしておきたい。なおこれ以降、物件の属性や項目に対応するカラムのことをフィールド、1物件を格納する行のことをレコードと呼ぶことにする。

1つのフィールド F に注目し、F の値ごとにレコードをグループ分けし、各グループごとに別のフィールド X (F 自身でもよい) の非空値の個数を数えたり (数値の場合) 合計や平均を求めることを**単純集計**という。このときの F を**分類フィールド**、X を**集計フィールド**と呼ぶことにする。また、このような単純集計を「F → X」と書くことにする。例えば (あまり意味のある集計ではないが) 図 2.5 のテーブルにおいて顧客名ごとに単価の合計を求める単純集計 (顧客名 → 単価) の結果は図 3.1 のような単純集計表 (Excel のピボットテーブルで作成) で表せる。

合計：単価	
顧客名	計
旭町工業	320
豊平商事	560
総計	880

(図 3.1) 単純集計表 (顧客名→単価)

2つのフィールド F, G に注目し, F と G の値の組み合わせごとにレコードをグループ分けし, 各グループごとに別のフィールド X (F または G 自身でもよい) の非空値の個数を数えたり (数値の場合) 合計や平均を求めることをクロス集計という。このときの F と G を分類フィールド, X を集計フィールドと呼ぶことにし, このようなクロス集計を「F×G→X」と書くことにする。例えば, 図 2.5 のテーブルにおいて NO と顧客名の組み合わせごとに単価の合計を求めるクロス集計 (NO×顧客名→単価) の結果は図 3.2 のようなクロス集計表 (Excel のピボットテーブルで作成) で表せる。

合計：単価	顧客名		
NO	旭町工業	豊平商事	総計
0001		560	560
0002	320		320
総計	320	560	880

(図 3.2) クロス集計表 (NO×顧客名→単価)

以上の2つの例は, 分類フィールドの取りうる値ごとにレコードをグループ分けしたが, 分類フィールドの値にグループ分けがあって, そのグループごとにレコードをグループ分けしたい場合がある。例えば, 単価が100円以下のレコードとそうでないレコードの数を数えるような場合である。このときは, 単価が100円以下ならば1, そうでなければ0を値としてとる新たなフィールド (例えば「低額」) を図 3.3 のように設ける。

NO	日付	顧客名	品名	単価	数量	低額
0001	2004/5/7	豊平商事	シャープペン	100	5	1
0001	2004/5/7	豊平商事	消しゴム	60	5	1
0001	2004/5/7	豊平商事	修正テープ	400	2	0
0002	2004/5/7	旭町工業	ボールペン黒	100	10	1
0002	2004/5/7	旭町工業	セロテープ大	220	5	0

(図 3.3) 補助フィールド「低額」を付加したテーブル

その上で単純集計「低額→顧客名」を行えば, 図 3.4 のような単純集計表が得られる。

データの個数：顧客名	
低額	計
	0 2
	1 3
総計	5

(図 3.4) 単純集計表 (低額→顧客名)

この「低額」フィールドのように、あるフィールドの値のグループを表すフィールドをここでは補助フィールドと呼ぶ。

4. Excel のデータ処理機能

この章では、第 3 章で述べた集計法を含む基本的なデータ処理を Excel で行う方法について整理する。説明のためのサンプルとして、あるシートの H9 から右下に、図 4.1 のように入力されたデータを考える。

	H	I	J	K	L	M
9	NO	日付	顧客名	品名	単価	数量
10	0001	2004/5/7	豊平商事	シャープペン	100	5
11	0001	2004/5/7	豊平商事	消しゴム	60	5
12	0001	2004/5/7	豊平商事	修正テープ	400	2
13	0002	2004/5/7	旭町工業	ボールペン黒	100	10
14	0002	2004/5/7	旭町工業	セロテープ大	220	5

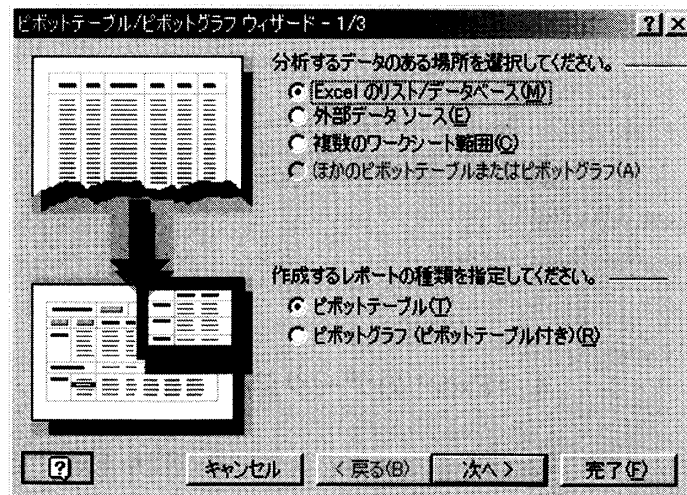
(図 4.1) 説明用サンプル

4.1. 集計機能

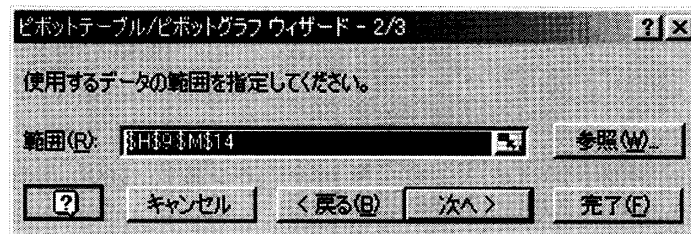
Excel の集計ツールとして、内訳を明示しながら単純集計を行うことが得意なアウトラインと、内訳表示はしないが単純集計やクロス集計を手軽に実行できるピボットテーブルの 2 つがある。ここでは後者について述べる。

〈例 4.1〉ピボットテーブルによるクロス集計 (NO×顧客名→単価) を同じワークシートの U9 より右下に行うには

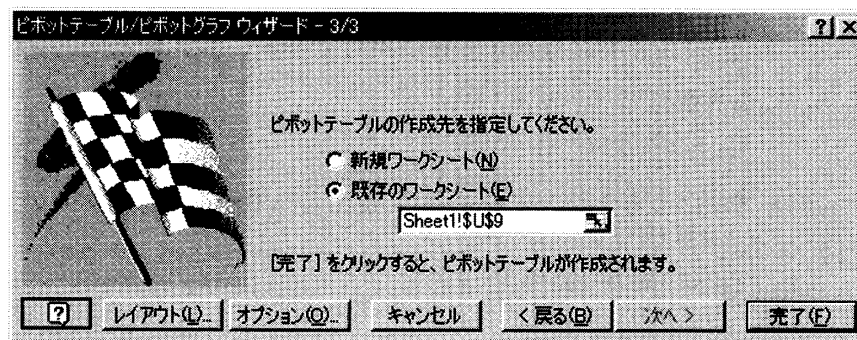
- ① テーブルの左上端セル (H9) をアクティブにする。
- ② 「データ」→「ピボットテーブルとピボットグラフレポート」の順にクリック。
- ③ 下図のままで「次へ」をクリック



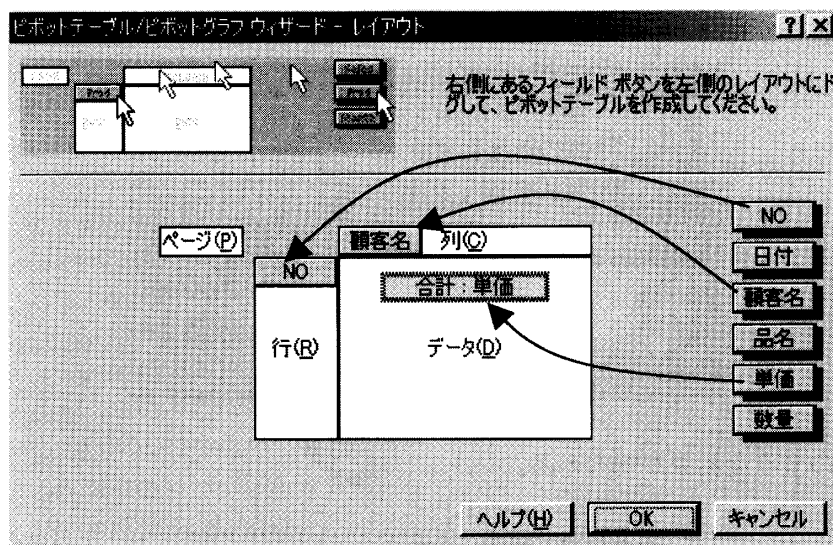
④下図のウィンドウではテーブルの範囲が合っていることを確かめて「次へ」をクリック。



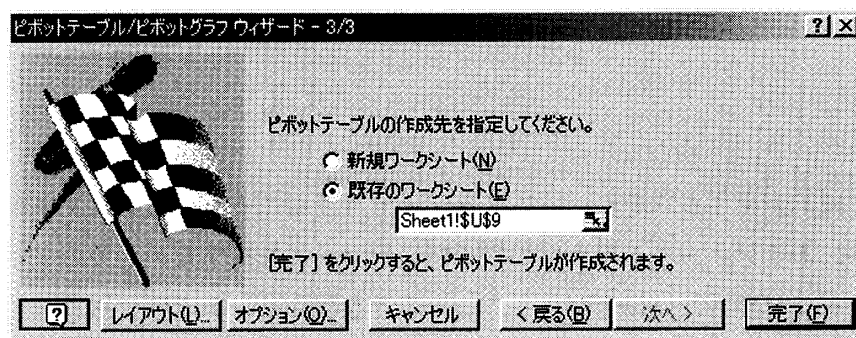
⑤「既存のワークシート」→出力先の左上端セル (U9) の順にクリックして下図のようになったら「レイアウト」をクリック。



⑥下図のように1つ目の分類フィールド (NO) を行エリアに、2つ目の分類フィールド (顧客名) を列エリアに、集計フィールド (単価) をデータエリアにそれぞれ配置 (ドラッグ) して「OK」をクリック。



⑦下図のウィンドウで「完了」をクリック。



以上の操作により、図 4.2 のようなピボットテーブルが指定した場所に作成される。

	U	V	W	X
9	合計：単価	顧客名		
10	NO	旭町工業	豊平商事	総計
11	0001		560	560
12	0002	320		320
13	総計	320	560	880

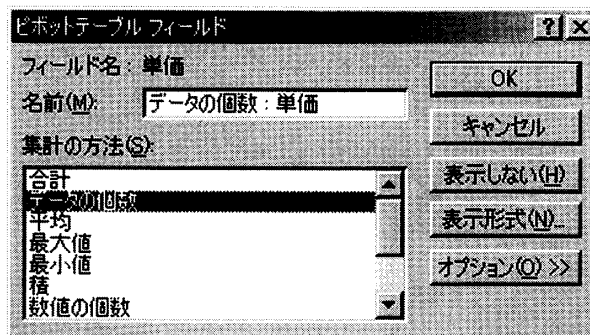
(図 4.2) ピボットテーブルによるクロス集計表

なお、ピボットテーブルは集計フィールドの値が数値の場合は合計を、文字列の場合は個数をそれぞれ自動的に計算する。集計方法を変更する（例えば合計から個数に変える）には次のようにすればよい。

<例 4.2> 例 4.1 で得たピボットテーブルの集計方法を合計から個数に変更するには

- ①ピボットテーブルの左上端セル（「合計：単価」とある U9）をダブルクリック。

②下図で「データの個数」をクリックして「OK」をクリック。



こうして図 4.3 のような結果が得られる。

データの個数: 単価	顧客名		
NO	旭町工業	豊平商事	総計
0001		3	3
0002	2		2
総計	2	3	5

(図 4.3) 集計法を変更したピボットテーブル

4.2. 検索機能

前節のように集計表を作る以前に、注目するフィールドがある条件を満たすようなレコードだけを抽出したり、あるフィールドの上位の値を持つレコードを探すため、そのフィールドのある順番でレコードを並べ替えたい場合がある。このような操作は、第5章で扱う非正規形データの集計時に必要になることもある。なぜなら、非正規形データではピボットテーブルによる集計が困難な場合があり、そのときは分類フィールドの値ごとにレコードを抽出または並べ替えをして(次節の subtotal 関数などで)レコード数を数え、手で集計表を作らなければならないからである。さらに正規形データにおいても、第3章で述べた補助フィールドを作る際に抽出または並べ替えを行う場合がある。そこで本稿では、抽出の方法として操作が簡単なオートフィルタと、操作はやや面倒だが汎用性が高いフィルタオプションについて述べ、次に並べ替えの方法に触れる。

オートフィルタ

オートフィルタを利用するには、抽出の対象となるテーブルをオートフィルタモードに変更する必要がある。

<例 4.3> 図 4.1 のテーブルをオートフィルタモードにするには

- ① テーブルの左上端セル (H9) をアクティブにする。
- ② 「データ」 → 「フィルタ」 → 「オートフィルタ」の順にクリック。

各フィールド名に [▼] ボタンが現れれば準備完了である。

〈例 4.4〉 NO フィールドの値が 0002 であるレコードだけを抽出するには

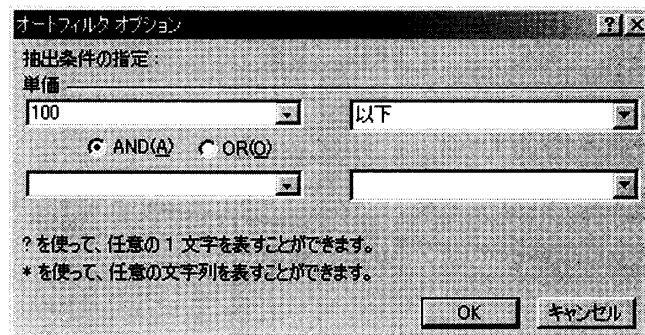
- ① NO フィールドの [▼] をクリック。
- ②プルダウンメニューの中の「0002」をクリック。

〈例 4.5〉 NO フィールドの条件をはずしてすべてのレコードを表示するには

- ① NO フィールドの [▼] (青色) をクリック。
- ②プルダウンメニューの中の「(すべて)」をクリック。

〈例 4.6〉 NO フィールドの値が 0002 で、かつ単価が 100 円以下のレコードを抽出するには

- ①例 4.4 の方法で NO が 0002 のレコードを抽出。
- ②単価フィールドの [▼] をクリック。
- ③プルダウンメニューの「(オプション)」をクリック。
- ④下図のように「100」を入力、「以下」を [▼] から選択して「OK」をクリック。



オートフィルタと次に述べるフィルタオプションは併用できないので、フィルタオプションを使う時は次の手順でオートフィルタモードを解除する。

〈例 4.7〉 オートフィルタモードを解除するには

- ①「データ」→「フィルタ」→「オートフィルタ」の順にクリック。

フィルタオプション

上記のようにオートフィルタはレコードの抽出を手軽に実現するが、1つの大きな弱点がある。それは、異なるフィールドの and (かつ) 条件は例 4.6 のように設定できても、or (または) 条件ができないことである。しかし非正規形データでは、繰返しフィールドが形式上異なるフィールドと見なされるので、繰返しフィールドに条件を設定するには異なるフィールドの or 条件を設定

できる必要がある。それも含めて、より柔軟な条件による抽出をするには多少面倒だがフィルタオプションを用いると良い。

フィルタオプションを利用するには、図4.4のように抽出条件を入力するためのエリアを用意する必要がある。

	H	I	J	K	L	M
9	NO	日付	顧客名	品名	単価	数量
10	0001	2004/5/7	豊平商事	シャープペン	100	5
11	0001	2004/5/7	豊平商事	消しゴム	60	5
12	0001	2004/5/7	豊平商事	修正テープ	400	2
13	0002	2004/5/7	旭町工業	ボールペン黒	100	10
14	0002	2004/5/7	旭町工業	セロテープ大	220	5
15						
16	NO	日付	顧客名	品名	単価	数量
17						
18						

← H9:M9の文字列をコピー

(図4.4) 16行目以降に抽出条件エリアを付加したレイアウト

<例4.8> NOフィールドの値が0002であるレコードだけを抽出するには

①下図のように抽出条件を入力。

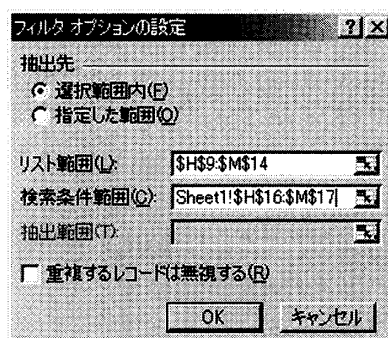
	H	I	J	K	L	M
9	NO	日付	顧客名	品名	単価	数量
10	0001	2004/5/7	豊平商事	シャープペン	100	5
11	0001	2004/5/7	豊平商事	消しゴム	60	5
12	0001	2004/5/7	豊平商事	修正テープ	400	2
13	0002	2004/5/7	旭町工業	ボールペン黒	100	10
14	0002	2004/5/7	旭町工業	セロテープ大	220	5
15						
16	NO	日付	顧客名	品名	単価	数量
17	0002					
18						

注：最初にシングルクォートを付けて '0002 と入力

②テーブルの左上端 (H9) をアクティブにする。

③「データ」→「フィルタ」→「フィルタオプションの設定」の順にクリック。

④リスト範囲がテーブルの範囲 (H9:M14) であることを確認し、検索条件範囲が条件用のフィールド名 (16行目) と条件式 (17行目) を含むような範囲 (下図) に指定して「OK」をクリック。



〈例 4.9〉 NO フィールドの条件をはずしてすべてのレコードを表示するには

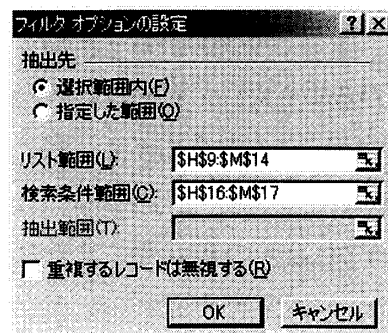
- ① 「データ」 → 「フィルタ」 → 「すべて表示」の順にクリック。

〈例 4.10〉 NO フィールドの値が 0002 で、かつ単価が 100 円以下のレコードを抽出するには

- ① 下図のように抽出条件を入力

	H	I	J	K	L	M
9	NO	日付	顧客名	品名	単価	数量
10	0001	2004/5/7	豊平商事	シャープペン	100	5
11	0001	2004/5/7	豊平商事	消しゴム	60	5
12	0001	2004/5/7	豊平商事	修正テープ	400	2
13	0002	2004/5/7	旭町工業	ボールペン黒	100	10
14	0002	2004/5/7	旭町工業	ゼロテープ大	220	5
15						
16	NO	日付	顧客名	品名	単価	数量
17	0002				<=100	
18						

- ② テーブルの左上端 (H9) をアクティブにする。
- ③ 「データ」 → 「フィルタ」 → 「フィルタオプションの設定」の順にクリック。
- ④ リスト範囲がテーブルの範囲 (H9:M14) であることを確認し、検索条件範囲が条件用のフィールド名 (16 行目) と条件式 (17 行目) を含むような範囲 (下図) に指定して「OK」をクリック。



例4.10の「単価」フィールドのような条件式を表すための演算子には以下のようなものがある。

<= (以下), >= (以上), > (より大きい), < (より小さい, 未満),
= (等しい), <> (等しくない)。

<例4.11> NOフィールドの値が0002か、または単価が100円以下のレコードを抽出するには

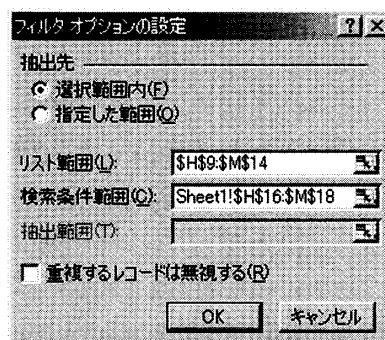
①下図のように抽出条件を入力

	H	I	J	K	L	M
9	NO	日付	顧客名	品名	単価	数量
10	0001	2004/5/7	豊平商事	シャープペン	100	5
11	0001	2004/5/7	豊平商事	消しゴム	60	5
12	0001	2004/5/7	豊平商事	修正テープ	400	2
13	0002	2004/5/7	旭町工業	ボールペン黒	100	10
14	0002	2004/5/7	旭町工業	セロテープ大	220	5
15						
16	NO	日付	顧客名	品名	単価	数量
17	0002					
18					<=100	

②テーブルの左上端 (H9) をアクティブにする。

③「データ」→「フィルタ」→「フィルタオプションの設定」の順にクリック。

④リスト範囲がテーブルの範囲 (H9:M14) であることを確認し、検索条件範囲が条件用のフィールド名 (16行目) と条件式 (17, 18行目) を含むような範囲 (下図) に指定して「OK」をクリック。



例4.10と例4.11を比較するとわかるように、フィルタオプションにおいてはandで結ばれる条件は同じ行に、orで結ばれる条件は異なる行に入力する。例えば、3つのフィールドX, Y, Zに関して図4.5のように入力される条件は

「XがC1」か、または「YがC2かつC3」か、または「XがC4かつZがC5」を満たすことを意味する。

X	Y	Y	Z
C1			
	C2	C3	
C4			C5

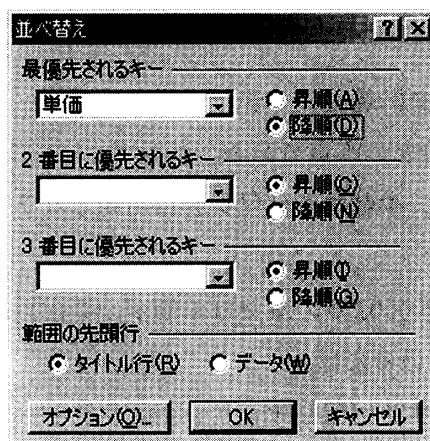
(図 4.5) 複雑な抽出条件

並べ替え³

あるフィールドの値が上位（または下位）のレコードを探すには並べ替え（ソート）の操作が有用である。

〈例 4.12〉 レコードをあるフィールド（例えば単価）の降順に並べ替えるには

- ① テーブルの左上端 (H9) をアクティブにする。
- ② 「データ」 → 「並べ替え」の順にクリック。
- ③ 下図のように最優先されるキーで大小比較するフィールド (単価) と「降順」を選択して「OK」をクリック。



〈例 4.13〉 レコードの順番を元に戻すには

- ① 「編集」 → 「元に戻す」の順にクリック。

なお、レコードの順番を元に戻す方法として例 4.13 はあまり好ましくない。なぜなら、並べ替えた後に他の処理を行うと、(順番を元に戻すため例 4.13 を繰り返すうち) それらの処理まで取り消されるからである。それを回避するには、テーブル作成時にレコードの通し番号を表すフィールドを設け、レコードの順番を元に戻す操作は通し番号フィールドの昇順に並べ替える操作で行

3 前項の実習でデータが条件付表示のままである場合は、例 4.9 の操作ですべて表示の状態に戻しておくこと。

えばよい。

4.3. 集計関数

Excelにはデータを集計するための関数が豊富に用意されているが、ここでは汎用性と柔軟性が高いと思われる3つの関数を取り上げる。

subtotal 関数

Excelの集計関数は、集計する対象や方法ごとに数多く用意されている。代表的なものにcount(数値データの個数を数える関数)、counta(非空値セルの個数を数える関数)、sum(数値データの合計を計算する関数)、average(数値データの平均を求める関数)などがある。そしてこれらの機能をすべて兼ね備えた関数としてsubtotal関数がある。その書式は以下のとおりである。

=subtotal(集計方法, 集計範囲)

但し、第1引数の集計方法には次のような意味を表す1から11までの値を指定する。

- 1：平均, 2：数値の個数, 3：非空値セルの個数, 4：最大値, 5：最小値, 6：積,
7：標本標準偏差, 8：母標準偏差, 9：合計, 10：標本分散, 11：母分散

subtotal関数の大きな特徴は、集計範囲内で表示されているデータだけが処理対象になる(つまり非表示データは処理されない)という点である。よって、4.2節のオートフィルタまたはフィルタオプションとの併用によって、柔軟な条件付集計が可能となる。

〈例4.14〉NOフィールドの値が0002か、または単価が100円以下のレコードの件数をJ20に、単価の合計をL20にそれぞれ求めるには

- ① J20に次のような式を入力する(但しJ10:J14は顧客名データがある範囲)。

=subtotal(3,J10:J14)

- ② L20に次のような式を入力する(但しL10:L14は単価データがある範囲)。

=subtotal(9,L10:L14)

- ③例4.11に従って条件に合うレコードを抽出する。(結果確認後は例4.9の方法で元に戻す。)

countif 関数, sumif 関数

上記のsubtotal関数は、非繰り返しフィールドのように集計対象が縦1列に並んでいる場合は適用できるが、繰り返しフィールドのように集計対象が縦横に並んでいる場合うまく集計できないことがある。その場合はcountif(条件を満たすセルの個数を数える関数)やsumif(条件を満たすセルに対応する数値データの合計を求める関数)で克服できることがある。まず、countif関

数の書式は次のとおりである。

=countif(集計範囲, 検索条件)

検索条件にはフィルタオプションで記入するような条件式を2つのダブルクォート (") で囲んで指定する。

<例 4.15> 顧客名データのある範囲 (J10:J14) で顧客名が「豊」で始まるセルの個数を J21 に求めるには

① J21 に次のような式を入力する。

=countif(J10:J14,"豊*")

一方, sumif 関数の書式は次のとおりである。

=sumif(条件範囲, 検索条件[, 集計範囲])

なお, [] は省略可能であることを意味し, 省略した場合は条件範囲が集計範囲となる。

<例 4.16> 顧客名が「豊」で始まるレコードの単価の合計を L21 に求めるには

① L21 に次のような式を入力する (但し J10:J14 は顧客名データが, L10:L14 は単価データがそれぞれある範囲)。

=sumif(J10:J14,"豊*",L10:L14)

これら countif 関数や sumif 関数も決して万能ではなく, 両者とも (その書式から) 分離した範囲の指定ができないという難点がある。また sumif 関数は, 条件範囲と集計範囲の行列数を等しくしないと正しい計算がなされないので注意を要する。

5. Excel による非正規形データの集計

第4章で述べた手法を用いて, 非正規形データを実際に集計する方法について検討する。前提として, 図5.1のようなサンプルデータがあるとしよう。このデータは, ある文房具店における売上記録を表すものとする。

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y		
1	NO	日付	顧客コード	取扱分類	商品コード	単価1	数量1	小計1	商品コード2	単価2	数量2	小計2	商品コード3	単価3	数量3	小計3	商品コード4	単価4	数量4	小計4	商品コード5	単価5	数量5	小計5	合計	
2	1	4/1	0009H	2	D0001A3	2000	50	120000	D0001A4	1000	50	50000	D0001B4	1500	50	80000									0	270000
3	2	4/1	0004A	2	A0002B5	350	20	7000																	0	7000
4	3	4/1	0003H	3	B0001B	430	5	2150	B0002	120	10	1200													0	3350
5	4	4/1	0006S	3	D0001B3	730	10	7300	A0001B5	390	30	11700	A0003B5	75	30	2250	C0001B4	100	20	2000	C0005	400	3	1200	24850	
6	5	4/2	0003H	3	B0005B	80	5	400																	0	400
7	6	4/2	0001D	1	A0001A4	400	100	40000	A0002A4	360	100	36000	A0003A4	80	100	8000	B0001L	550	30	16500					0	101500
8	7	4/2	0004A	2	S0003	100	50	5000	C0004	50	30	1500													0	7500
9	8	4/3	0005T	2	A0002B5	350	30	10500	A0003B5	75	50	3750	D0001B5	750	20	15000									0	30000
10	9	4/3	0006S	3	B0001B	430	5	2150																	0	2150
11	10	4/5	0002T	1	C0001B4	100	30	3000	C0001R	100	30	3000	C0001BL	100	30	3000									0	9000
12	11	4/5	0001D	1	C0002B4	150	50	7500	C0002R	150	50	7500	C0002BL	150	50	7500									0	22500
13	12	4/5	0006H	2	B0004L	5800	3	20400	B0005L	550	6	3300													0	23700
14	13	4/5	0007B	3	B0003L	220	10	2200																	0	2200
15	14	4/6	0007B	3	B0005S	80	5	400																	0	400
16	15	4/6	0009W	3	B0003L	220	5	1100	B0004S	350	3	1050													0	2150
17	16	4/7	0009W	3	C0002B4	150	20	3000	C0002R	150	10	1500	C0002BL	150	10	1500									0	6000
18	17	4/7	0006H	2	D0001B4	1500	120	180000																	0	180000
19	18	4/7	0005T	2	C0001B4	100	50	5000	C0001R	100	50	5000													0	10000
20																										

(図 5.1) サンプルデータ (非正規形)

図 5.1 において, 各レコードは 1 件分の売上を表し, 「商品コード」「単価」「数量」「小計」はそ

の売上の内訳を表す繰り返しフィールドである。繰り返しフィールドの名前の末尾には、内訳番号を表す数字が付加されている。ここでは内訳の最大件数を5と仮定する。一方、「NO」「日付」「顧客コード」「規模分類」を非繰り返しフィールドと見なす。顧客情報と商品情報の詳細は図5.2のとおりとする。

顧客コード	顧客名	規模分類	商品コード	商品名	単価	分類
0001D	大北海建設	1	A0001A4	リング式ファイルA 4	400	ファイル
0002T	豊平商事	1	A0001B5	リング式ファイルB 5	390	ファイル
0003H	平岸不動産	3	A0002A4	Z式ファイルA 4	360	ファイル
0004A	旭町工業	2	A0002B5	Z式ファイルB 5	350	ファイル
0005T	月寒食品	2	A0003A4	クリアホルダーA 4	90	ファイル
0006H	東札幌印刷	2	A0003B5	クリアホルダーB 5	75	ファイル
0007S	佐藤商店	3	B0001L	ゼムクリップ大 (1000)	550	事務用品
0008S	鈴木会計事務所	3	B0001S	ゼムクリップ小 (1000)	430	事務用品
0009W	渡辺運輸	3	B0002	スティックのり	120	事務用品
			B0003L	セロテープ大	220	事務用品
			B0003S	セロテープ小	120	事務用品
			B0004L	ホッチキス大	6800	事務用品
			B0004S	ホッチキス小	350	事務用品
			B0005L	ホッチキス針大 (1000)	550	事務用品
			B0005S	ホッチキス針小 (1000)	80	事務用品
			C0001BK	ボールペン黒	100	筆記用具
			C0001R	ボールペン赤	100	筆記用具
			C0001BL	ボールペン青	100	筆記用具
			C0002BK	油性マーカー黒	150	筆記用具
			C0002R	油性マーカー赤	150	筆記用具
			C0002BL	油性マーカー青	150	筆記用具
			C0003	シャープペン	100	筆記用具
			C0004	消しゴム	60	筆記用具
			C0005	修正テープ	400	筆記用具
			D0001A3	PPC用紙A 3 (500)	2000	OA用紙
			D0001A4	PPC用紙A 4 (500)	1000	OA用紙
			D0001B4	PPC用紙B 4 (500)	1500	OA用紙
			D0001B5	PPC用紙B 5 (500)	750	OA用紙

(図5.2) 顧客情報(左)と商品情報(右)

顧客情報の中で、分類規模は1が大企業、2が中小企業、3が零細企業を表すとする。また、商品コードの先頭文字は商品の分類に対応しているとする。

分析の下準備として次の作業を行い、図5.3のようなレイアウトを作成する。

- ① Z列に各レコードの商品コードの個数を表すフィールド「品数」を作る。具体的にはZ2に
- $$= \text{subtotal}(3, E2, I2, M2, Q2, U2)$$

を入力し、それをZ3:Z19に複写する。

② 1行目のフィールド名を縦書きにして列幅を詰める。

③ フィルタオプションのための検索条件入力範囲をテーブルの上にする。条件入力のための空行は少なくとも5行 (=内訳件数の最大数) 設ける (図5.3では6行設けた)。

この章では、図5.3のような作業レイアウトを用いた単純集計とクロス集計の方法を、さまざまな場合について考察する。

	A N O	B 日 付	C 顧 客 コ ー ド	D 規 模 分 類	E 商 品 コ ー ド 1	F 単 価 1	G 数 量 1	H 小 計 1	I 商 品 コ ー ド 2	J 単 価 2	K 数 量 2	L 小 計 2	M 商 品 コ ー ド 3	N 単 価 3	O 数 量 3	P 小 計 3	Q 商 品 コ ー ド 4	R 単 価 4	S 数 量 4	T 小 計 4	U 商 品 コ ー ド 5	V 単 価 5	W 数 量 5	X 小 計 5	Y 合 計	Z 品 数	
1																											
2																											
3																											
4																											
5																											
6																											
7																											
	N O	日 付	顧 客 コ ー ド	規 模 分 類	商 品 コ ー ド 1	単 価 1	数 量 1	小 計 1	商 品 コ ー ド 2	単 価 2	数 量 2	小 計 2	商 品 コ ー ド 3	単 価 3	数 量 3	小 計 3	商 品 コ ー ド 4	単 価 4	数 量 4	小 計 4	商 品 コ ー ド 5	単 価 5	数 量 5	小 計 5	合 計	品 数	
9	1	4/1	0006H	2	D0001A3	2000	60	120000	D0001A4	1000	60	60000	D0001B4	1500	60	90000									0	270000	3
10	2	4/1	0004A	2	A0002B5	350	20	7000				0				0								0	7000	1	
11	3	4/1	0003H	3	B0001S	430	5	2150	B0002	120	10	1200				0								0	3350	2	
12	4	4/1	0008S	3	D0001B5	750	10	7500	A0001B5	390	30	11700	A0003B5	75	30	2250	C0001BK	100	20	2000	C0005	400	3	1200	24650	5	
13	5	4/2	0003H	3	B0005S	80	5	400				0				0								0	400	1	
14	6	4/2	0001D	1	A0001A4	400	100	40000	A0002A4	360	100	36000	A0003A4	90	100	9000	B0001L	550	30	16500					0	101500	4
15	7	4/2	0004A	2	C0003	100	60	6000	C0004	80	30	1800				0								0	7800	2	
16	8	4/3	0005T	2	A0002B5	350	30	10500	A0003B5	75	60	4500	D0001B5	750	20	15000								0	30000	3	
17	9	4/3	0008S	3	B0001S	430	5	2150				0				0								0	2150	1	
18	10	4/5	0002T	1	C0001BK	100	30	3000	C0001R	100	30	3000	C0001BL	100	30	3000								0	9000	3	
19	11	4/5	0001D	1	C0002BK	150	50	7500	C0002R	150	50	7500	C0002BL	150	50	7500								0	22500	3	
20	12	4/5	0006H	2	B0004L	6800	3	20400	B0005L	550	6	3300				0								0	23700	2	
21	13	4/5	0007S	3	B0003L	220	10	2200				0				0								0	2200	1	
22	14	4/6	0007S	3	B0005S	80	5	400				0				0								0	400	1	
23	15	4/6	0009W	3	B0003L	220	5	1100	B0004S	350	3	1050				0								0	2150	2	
24	16	4/7	0009W	3	C0002BK	150	20	3000	C0002R	150	10	1500	C0002BL	150	10	1500								0	6000	3	
25	17	4/7	0006H	2	D0001B4	1500	120	180000				0				0								0	180000	1	
26	18	4/7	0005T	2	C0001BK	100	50	5000	C0001R	100	50	5000				0								0	10000	2	

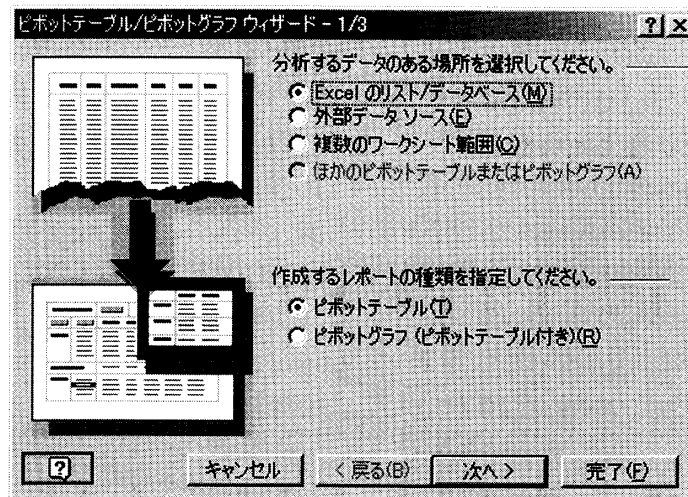
(図5.3) 作業用レイアウト

5.1. 非繰り返しフィールド→非繰り返しフィールド

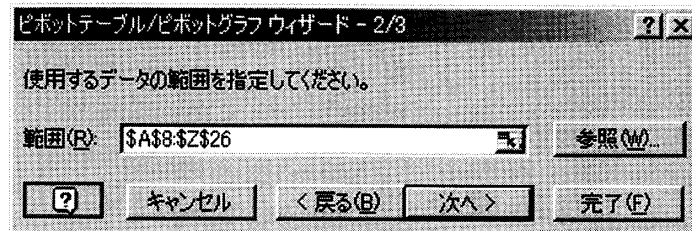
分類フィールド,集計フィールドともに非繰り返しフィールドである単純集計は、ピボットテーブルを用いて簡単に実行できる。

<例5.1> 日付ごとの顧客数を新たなワークシートに集計するには

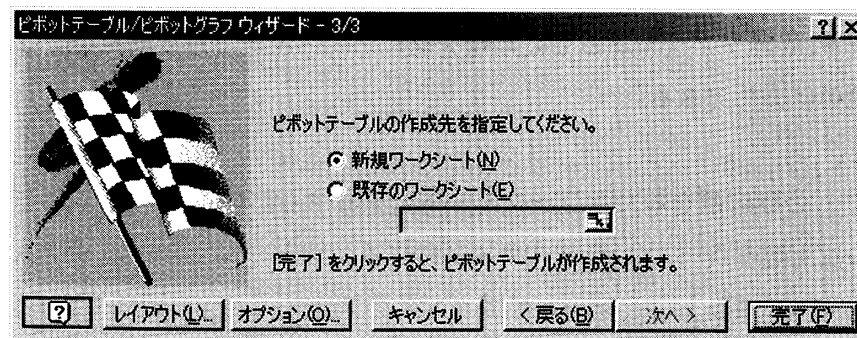
- ① テーブルの左上端セル (A8) をアクティブにする。
- ② 「データ」→「ピボットテーブルとピボットグラフレポート」の順にクリック。
- ③ 下図であることを確認して「次へ」をクリック。



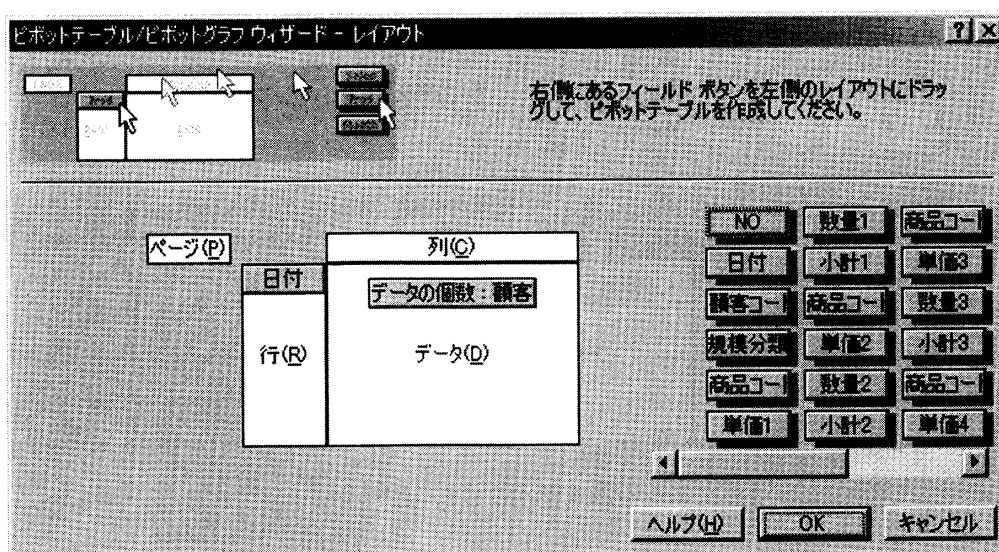
④下図のようにデータ範囲が正しいことを確認して「次へ」をクリック。



⑤下図では「新規ワークシート」を選択して「レイアウト」をクリック。



⑥レイアウトを下図のように設定して「OK」をクリック。



⑦再び手順⑤のウィンドウで「完了」をクリック。

以上の操作により、図 5.4 のようなピボットテーブルが得られる。

	A	B
3	データの個数: 顧客コード	
4	日付	計
5		4/1 4
6		4/2 3
7		4/3 2
8		4/5 4
9		4/6 2
10		4/7 3
11	総計	18

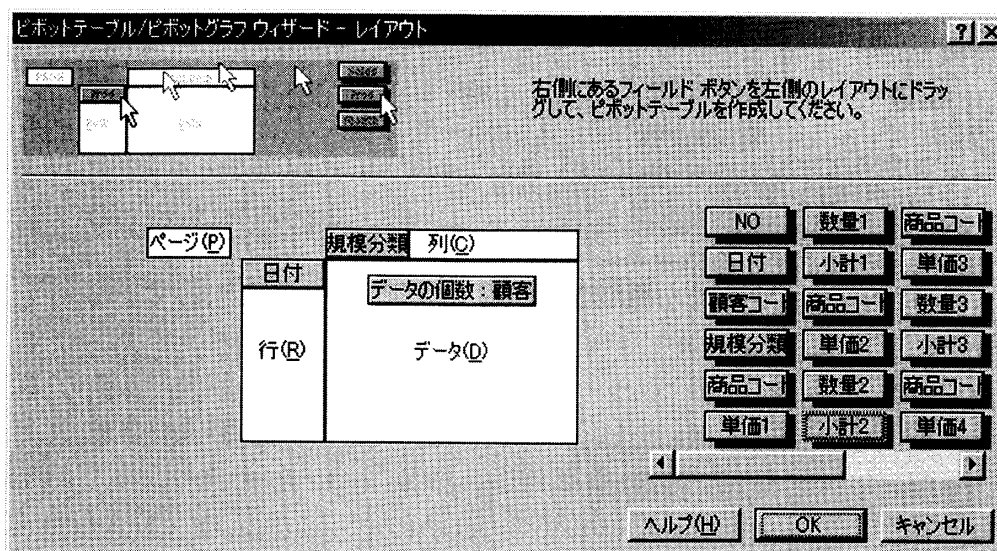
(図 5.4) 例 5.1 の実行結果

5.2. 非繰り返しフィールド×非繰り返しフィールド→非繰り返しフィールド

この場合のクロス集計もピボットテーブルで簡単にできる。例 5.1 と同様の方法で新たにピボットテーブルを作ることもできるが、ここでは既にあるピボットテーブルに修正を加えてクロス集計表を作ってみよう。

〈例 5.2〉 日付と規模分類の組み合わせごとの顧客数を求めるには

- ①ピボットテーブル内 (A3:B11) の任意のセルを右クリック。
- ②「ウィザード」→「レイアウト」の順にクリック。
- ③レイアウトを下図のように修正して「OK」をクリック。



④ 「完了」をクリック。

以上の操作により、図 5.5 のようなピボットテーブルが得られる。

	A	B	C	D	E
3	データの個数: 顧客コード	規模分類			
4	日付	1	2	3	総計
5	4/1		2	2	4
6	4/2	1	1	1	3
7	4/3		1	1	2
8	4/5	2	1	1	4
9	4/6			2	2
10	4/7		2	1	3
11	総計	3	7	8	18

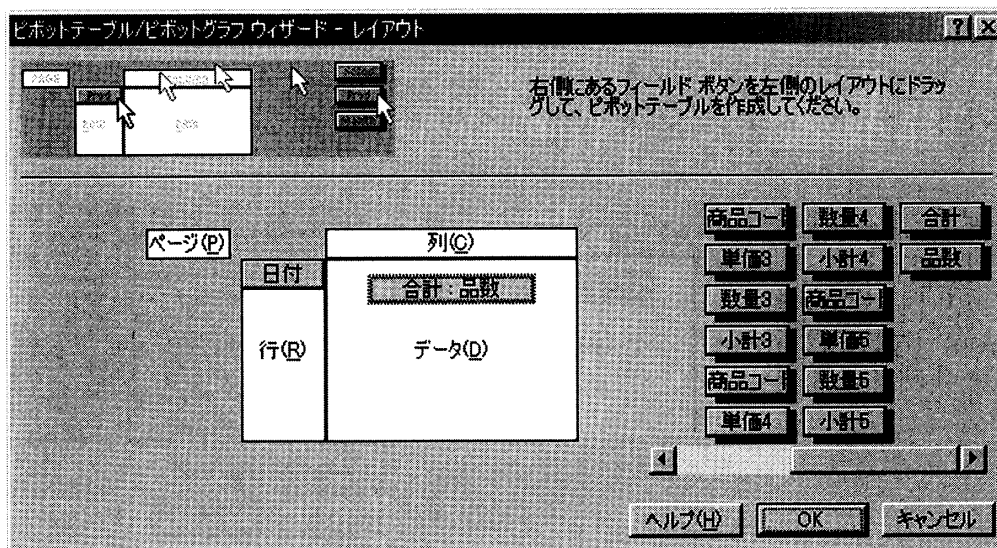
(図 5.5) 例 5.2 の実行結果

5.3. 非繰り返しフィールド→繰り返しフィールド

このような単純集計も、各レコードごとの集計フィールド (図 5.3 では Y 列の「合計」と Z 列の「品数」)があれば、ピボットテーブルで簡単に実行できる。ここでも、例 5.2 で得られたピボットテーブルを再利用する。

〈例 5.3〉 日付ごとに商品コードの個数を求めるには

- ①ピボットテーブル内 (A3:E11) の任意のセルを右クリック。
- ②「ウィザード」→「レイアウト」の順にクリック。
- ③列エリアとデータエリアにあるフィールド名をエリア外に移動 (ドラッグ) して消去し、「品数」フィールドをデータエリアに移動し、下図のようになったら「OK」をクリック。



④ 「完了」をクリック。

以上の操作により、図 5.6 のようなピボットテーブルが得られる。なお、日付ごとの売上合計も、例 5.3 の手順③で「品数」の代わりに「合計」を移動することにより得られる。

	A	B
3	合計:品数	
4	日付	計
5	4/1	11
6	4/2	7
7	4/3	4
8	4/5	9
9	4/6	3
10	4/7	6
11	総計	40

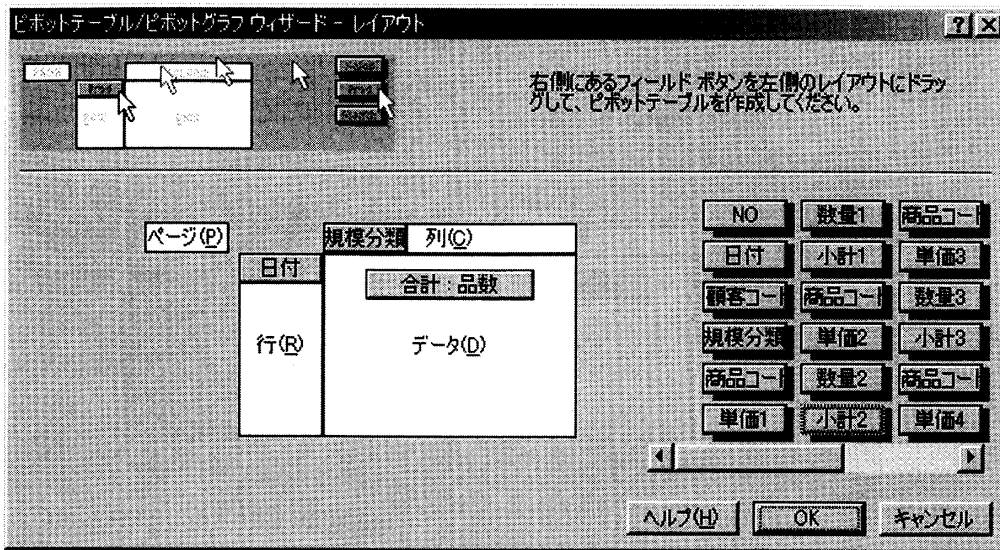
(図 5.6) 例 5.3 の実行結果

5.4. 非繰り返しフィールド×非繰り返しフィールド→繰り返しフィールド

このようなクロス集計もピボットテーブルで簡単にできる。

〈例 5.4〉 日付と規模分類の組み合わせごとの商品コードの個数を求めるには

- ①ピボットテーブル内 (A3:B11) の任意のセルを右クリック。
- ②「ウィザード」→「レイアウト」の順にクリック。
- ③レイアウトを下図のように修正して「OK」をクリック。



④「完了」をクリック。

以上の操作により、図5.7のようなピボットテーブルが得られる。なお、日付と規模分類の組み合わせごとの売上合計も、例5.4の手順③でさらにデータエリアのフィールドを「合計」に変更することにより得られる。

	A	B	C	D	E
3	合計：品数	規模分類			
4	日付	1	2	3	総計
5	4/1		4	7	11
6	4/2	4	2	1	7
7	4/3		3	1	4
8	4/5	6	2	1	9
9	4/6			3	3
10	4/7		3	3	6
11	総計	10	14	16	40

(図5.7) 例5.4の実行結果

5.5. 繰り返しフィールド→非繰り返しフィールド

前節までのように、分類フィールドがすべて非繰り返しフィールドの単純およびクロス集計はピボットテーブルで簡単に実行できたが、分類フィールドが繰り返しフィールドになると集計が急に難しくなる。

まず、繰り返しフィールドによる非繰り返しフィールドの単純集計について検討しよう。集計の際の要件として、分類フィールドにおける分類はレコード集合を互いに素に分割するものとする。例えば、「商品分類Cを購入したレコードとそうでないレコード」という分類は上記の要件を満たすが、「商品分類Aを購入したレコード、Bを購入したレコード、……」という分類は一般に互いに素とならない。上記の要件を満たす分類は、フィルタオプションを用いて補助フィールド

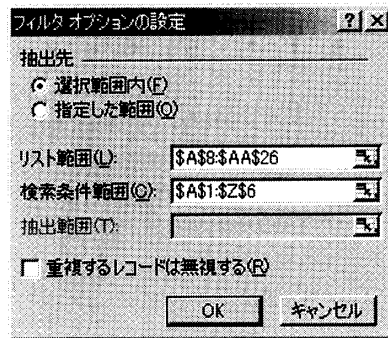
に記入しておくとい。

<例 5.5> 商品分類Cを購入したか否かにより顧客数を数えるには

- ①補助フィールド名として例えば「C購入」をAA8に記入する。
- ②商品コードに関する条件を以下のように記入する。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	NO	日付	顧客コード	規模分類	商品コード1	単価1	数量1	小計1	商品コード2	単価2	数量2	小計2	商品コード3	単価3	数量3	小計3	商品コード4	単価4	数量4	小計4	商品コード5	単価5	数量5	小計5	合計	品数
1																										
2					C*																					
3									C*																	
4										C*																
5													C*													
6																					C*					
7																										

- ③テーブルの左上端セル (A8) をクリック。
- ④「データ」→「フィルタ」→「フィルタオプションの設定」の順にクリック。
- ⑤下図のように設定して「OK」をクリック。



- ⑥抽出されたレコードの「C購入」フィールドに条件合致を表す値 (例えば1) を下図のように入力。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
	NO	日付	顧客コード	規模分類	商品コード1	単価1	数量1	小計1	商品コード2	単価2	数量2	小計2	商品コード3	単価3	数量3	小計3	商品コード4	単価4	数量4	小計4	商品コード5	単価5	数量5	小計5	合計	品数	C購入	
8																												
12	4	4/1	0008S	3	D0001B5	750	10	7500	A0001B5	390	30	11700	A0003B5	75	30	2250	C0001BK	100	20	2000	C0005	400	3	1200	24650	5	1	
15	7	4/2	0004A	2	C0003	100	60	6000	C0004	60	30	1800																
18	10	4/5	0002T	1	C0001BK	100	30	3000	C0001R	100	30	3000	C0001BL	100	30	3000												
19	11	4/5	0001D	1	C0002BK	150	50	7500	C0002R	150	50	7500	C0002BL	150	50	7500												
24	16	4/7	0009W	3	C0002BK	150	20	3000	C0002R	150	10	1500	C0002BL	150	10	1500												
26	18	4/7	0005T	2	C0001BK	100	50	5000	C0001R	100	50	5000																

- ⑦「データ」→「フィルタ」→「すべて表示」の順にクリック。
- ⑧ピボットテーブルで単純集計表「C購入→顧客コード」を作成。

以上の操作により、図 5.8 のようなピボットテーブルが得られる。

データの個数：顧客コード	
C購入	計
	1
(空白)	12
総計	18

(図5.8) 例5.5の実行結果

5.6. 繰り返しフィールド×繰り返しフィールド→非繰り返しフィールド

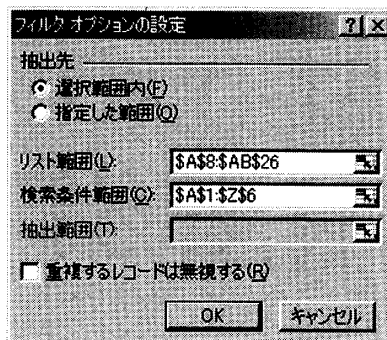
この場合も、互いに素な分類を表す補助フィールドとピボットテーブルを併用することにより集計可能である。

<例5.6> 「商品分類Cを購入したか否か」と「数量が50以上か否か」の組み合わせごとに顧客数を数えるには

- ①新たな補助フィールド名として例えば「大量購入」をAB8に記入する。
- ②商品コードに関する条件を以下のように記入する。

	A NO	B 日付	C 顧客 コード	D 規模 分類	E 商品 コード 1	F 単 価 1	G 数 量 1	H 小 計 1	I 商品 コード 2	J 単 価 2	K 数 量 2	L 小 計 2	M 商品 コード 3	N 単 価 3	O 数 量 3	P 小 計 3	Q 商品 コード 4	R 単 価 4	S 数 量 4	T 小 計 4	U 商品 コード 5	V 単 価 5	W 数 量 5	X 小 計 5	Y 合 計	Z 品 数
1																										
2							>=50																			
3											>=50															
4															>=50											
5																				>=50						
6																								>=50		
7																										

- ③テーブルの左上端セル (A8) をクリック。
- ④「データ」→「フィルタ」→「フィルタオプションの設定」の順にクリック。
- ⑤下図のように設定して「OK」をクリック。



- ⑥抽出されたレコードの「大量購入」フィールドに条件合致を表す値 (例えば1) を下図のように入力。

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
	商品 コード 4	単 価 4	数 量 4	小 計 4	商品 コード 5	単 価 5	数 量 5	小 計 5	合 計	品 数	C 購 入	大 量 購 入
8												
9				0				0	270000	3		1
14	B0001L	550	30	16500				0	101500	4		1
15				0				0	7800	2	1	1
16				0				0	30000	3		1
19				0				0	22500	3	1	1
25				0				0	180000	1		1
26				0				0	10000	2	1	1

- ⑦「データ」→「フィルタ」→「すべて表示」の順にクリック。
- ⑧ピボットテーブルでクロス集計表「C購入×大量購入→顧客コード」を作成。

以上の操作により図 5.9 のようなピボットテーブルが得られる。

データの個数：顧客コード		大量購入	
C購入		1 (空白)	総計
	1	3	3
(空白)		4	8
総計		7	11
			6
			12
			18

(図 5.9) 例 5.6 の実行結果

5.7. 繰り返しフィールド→繰り返しフィールド

このような集計を非正規形データからピボットテーブルを用いて直接行うことは、もはや不可能である。なぜならば、ピボットテーブルを利用するには分類フィールドの値と集計フィールドの値がそれぞれ 1 列に並んでいる必要があるからである。しかし、この場合の単純集計に限っては、countif 関数や sumif 関数を用いて手動で集計表を作ることが可能である。

ここでは、商品分類 (A~D) ごとの商品コードのカウントおよび売上の集計を試みる。その準備として、未使用のワークシートに図 5.10 のようなレイアウトを作成し、そのシート名を「商品分類別」とする。

	A	B	C
1		件数	売上
2	A		
3	B		
4	C		
5	D		

(図 5.10) 集計用レイアウト

なお、非正規形データ (図 5.3) があるシートの名前は「売上」とする。

<例 5.7.1> 商品分類別に商品コードの個数を「商品分類別」シートの B2:B5 に求めるには

- ① B2に次のような式を入力する(但し 売上!\$E\$9:\$U\$26 はすべての商品コードの値を含む範囲)。

=countif(売上!\$E\$9:\$U\$26, 商品分類別!A2 & "*"")

- ② B2の式を B3:B5に複写する。

手順①の式の中で、「商品分類別!A2 & "*"」は「商品分類別シートのA2の文字列に*を付加した条件式(B2では"A*")を表す。

<例5.7.2> 商品分類別に売上合計を「商品分類別」シートのC2:C5に求めるには

- ① C2に次のような式を入力する(但し 売上!\$H\$9:\$X\$26 はすべての小計の値を含む範囲)。

=sumif(売上!\$E\$9:\$U\$26,商品分類別!A2 & "*" ,売上!\$H\$9:\$X\$26)

- ② C2の式を C3:C5に複写する。

例5.7.1と例5.7.2の実行結果は図5.11のとおりである。

	A	B	C
1		件数	売上
2	A	8	120950
3	B	11	50850
4	C	15	58500
5	D	6	472500

(図5.11) 例5.7.1と例5.7.2の実行結果

5.8. 繰り返しフィールド×繰り返しフィールド→繰り返しフィールド

この場合は、もはやピボットテーブルはおろか countif 関数や sumif 関数による集計も不可能である。なぜならば、2つの関数とも分類フィールドを1つしか指定できないからである。しかし、非正規形データと同等の情報を持つ第1正規形データが図5.12のように得られていれば、ピボットテーブルによる集計が可能になる。

	A	B	C	D	E	F	G	H	I
1	NO	日付	顧客コード	規模分類	商品コード1	商品分類	単価1	数量1	大量購入
2	1	4/1	0006H	2	D0001A3	D	2000	60	1
3	1	4/1	0006H	2	D0001A4	D	1000	60	1
4	1	4/1	0006H	2	D0001B4	D	1500	60	1
5	2	4/1	0004A	2	A0002B5	A	350	20	0
6	3	4/1	0003H	3	B0001S	B	430	5	0
7	3	4/1	0003H	3	B0002	B	120	10	0
8	4	4/1	0008S	3	D0001B5	D	750	10	0
9	4	4/1	0008S	3	A0001B5	A	390	30	0
10	4	4/1	0008S	3	A0003B5	A	75	30	0
11	4	4/1	0008S	3	C0001BK	C	100	20	0
12	4	4/1	0008S	3	C0005	C	400	3	0
13	5	4/2	0003H	3	B0005S	B	80	5	0
14	6	4/2	0001D	1	A0001A4	A	400	100	1
15	6	4/2	0001D	1	A0002A4	A	360	100	1
16	6	4/2	0001D	1	A0003A4	A	90	100	1
17	6	4/2	0001D	1	B0001L	B	550	30	0
18	7	4/2	0004A	2	C0003	C	100	60	1
19	7	4/2	0004A	2	C0004	C	60	30	0
20	8	4/3	0005T	2	A0002B5	A	350	30	0
21	8	4/3	0005T	2	A0003B5	A	75	60	1
22	8	4/3	0005T	2	D0001B5	D	750	20	0
23	9	4/3	0008S	3	B0001S	B	430	5	0
24	10	4/5	0002T	1	C0001BK	C	100	30	0
25	10	4/5	0002T	1	C0001R	C	100	30	0
26	10	4/5	0002T	1	C0001BL	C	100	30	0
27	11	4/5	0001D	1	C0002BK	C	150	50	1
28	11	4/5	0001D	1	C0002R	C	150	50	1
29	11	4/5	0001D	1	C0002BL	C	150	50	1
30	12	4/5	0006H	2	B0004L	B	6800	3	0
31	12	4/5	0006H	2	B0005L	B	550	6	0
32	13	4/5	0007S	3	B0003L	B	220	10	0
33	14	4/6	0007S	3	B0005S	B	80	5	0
34	15	4/6	0009W	3	B0003L	B	220	5	0
35	15	4/6	0009W	3	B0004S	B	350	3	0
36	16	4/7	0009W	3	C0002BK	C	150	20	0
37	16	4/7	0009W	3	C0002R	C	150	10	0
38	16	4/7	0009W	3	C0002BL	C	150	10	0
39	17	4/7	0006H	2	D0001B4	D	1500	120	1
40	18	4/7	0005T	2	C0001BK	C	100	50	1
41	18	4/7	0005T	2	C0001R	C	100	50	1

(図 5.12) 第 1 正規形データ

但し、F 列は商品分類、I 列は数量が 50 以上を表す補助フィールドで、オートフィルタなどを用いて作成することができる。

〈例 5.8〉商品分類と「数量が 50 以上か否か」の組み合わせごとに商品コードの個数を数えるには

- ①図 5.12 のテーブルを用いてピボットテーブルでクロス集計「商品分類×大量購入→商品コード」を行う。

この操作により、図 5.13 のようなピボットテーブルが得られる。

データの個数：商品コード1	大量購入		
商品分類	0	1	総計
A	4	4	8
B	11		11
C	9	6	15
D	2	4	6
総計	26	14	40

(図 5.13) 例 5.8 の実行結果

以上から予想されるように、集計フィールドが繰り返しフィールドの場合は第1正規形データに直接ピボットテーブルを適用することにより、簡単に集計表を作成することができる。5.3節、5.4節、5.7節で行った集計作業は、いずれも非正規形データより第1正規形データを用いて行う方が簡単である。

5.9. 非繰り返しフィールド×繰り返しフィールド→非繰り返しフィールド

繰り返しフィールドの分類フィールドについては5.5節のように補助フィールドを設ければ、非正規形データに直接ピボットテーブルを適用して集計することができる。

〈例 5.9〉 規模分類と「商品分類Cを購入したか否か」の組み合わせごとに顧客数を求めるには

- ①例 5.5 のように、商品分類Cを購入したレコードのみ値1を持つ補助フィールド「C購入」を作成する。
- ②ピボットテーブルでクロス集計「規模分類×C購入→顧客コード」を実行する。

以上の結果、図 5.14 のようなピボットテーブルが得られる。

データの個数：顧客コード	C購入		
規模分類	1 (空白)		総計
1	2	1	3
2	2	5	7
3	2	6	8
総計	6	12	18

(図 5.14) 例 5.9 の実行結果

5.10. 非繰り返しフィールド×繰り返しフィールド→繰り返しフィールド

この場合は、分類フィールドである非繰り返しフィールドの並べ替えと5.7節の方法を組み合わせることで、非正規形データに対する集計が可能である。

ここでは、規模分類（非繰り返しフィールド）と商品分類（繰り返しフィールド）の組み合わせごとに商品コードの個数および売上合計を求める。その準備として、5.7節で作成した「商品分類別」シートに図 5.15 のようなレイアウトを追加する。

	A	B	C	D
9	件数	1	2	3
10	A			
11	B			
12	C			
13	D			
14				
15	売上	1	2	3
16	A			
17	B			
18	C			
19	D			

(図 5.15) 集計用レイアウト (追加分)

さらに、非正規形データのテーブル(「売上」シート内)を、図 5.16 のように規模分類の昇順で並べ替えておく(A8 をアクティブにして「データ」→「並べ替え」の順にクリックし、最優先されるキーを「規模分類」にすればよい)。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
	NO	日付	顧客コード	規模分類	商品コード1	単価1	数量1	小計1	商品コード2	単価2	数量2	小計2	商品コード3	単価3	数量3	小計3	商品コード4	単価4	数量4	小計4	商品コード5	単価5	数量5	小計5	合計	品数	
8																											
9	6	4/2	0001D	1	A0001A4	400	100	40000	A0002A4	360	100	36000	A0003A4	90	100	9000	B0001L	550	30	16500					0	101500	4
10	10	4/5	0002T	1	C0001BK	100	30	3000	C0001R	100	30	3000	C0001BL	100	30	3000									0	9000	3
11	11	4/5	0001D	1	C0002BK	150	50	7500	C0002R	150	50	7500	C0002BL	150	50	7500									0	22500	3
12	1	4/1	0006H	2	D0001A3	2000	60	120000	D0001A4	1000	60	60000	D0001B4	1500	60	90000									0	270000	3
13	2	4/1	0004A	2	A0002B5	350	20	7000				0				0								0	7000	1	
14	7	4/2	0004A	2	C0003	100	60	6000	C0004	60	30	1800				0								0	7800	2	
15	8	4/3	0005T	2	A0002B5	350	30	10500	A0003B5	75	60	4500	D0001B5	750	20	15000								0	30000	3	
16	12	4/5	0006H	2	B0004L	6800	3	20400	B0005L	550	6	3300				0								0	23700	2	
17	17	4/7	0006H	2	D0001B4	1500	120	180000				0				0								0	180000	1	
18	18	4/7	0005T	2	C0001BK	100	50	5000	C0001R	100	50	5000				0								0	10000	2	
19	3	4/1	0003H	3	B0001S	430	5	2150	B0002	120	10	1200				0								0	3350	2	
20	4	4/1	0008S	3	D0001B5	750	10	7500	A0001B5	390	30	11700	A0003B5	75	30	2250	C0001BK	100	20	2000	C0005	400	3	1200	24650	5	
21	5	4/2	0003H	3	B0005S	80	5	400				0				0								0	400	1	
22	9	4/3	0008S	3	B0001S	430	5	2150				0				0								0	2150	1	
23	13	4/5	0007S	3	B0003L	220	10	2200				0				0								0	2200	1	
24	14	4/6	0007S	3	B0005S	80	5	400				0				0								0	400	1	
25	15	4/6	0009W	3	B0003L	220	5	1100	B0004S	350	3	1050				0								0	2150	2	
26	16	4/7	0009W	3	C0002BK	150	20	3000	C0002R	150	10	1500	C0002BL	150	10	1500								0	6000	3	

(図 5.16) 規模分類で並べ替えた非正規形データ

<例 5.10.1> 規模分類と商品分類の組み合わせごとに商品コードの個数を求めるには

- ① 「商品分類別」シートの B10 に次のような式を入力する (但し 売上!\$E\$9:\$U\$11 は規模分類の値が 1 であるレコードのすべての商品コードを含む範囲)。

$$=countif(売上!E9:U11, 商品分類別!A10 & "*")$$

- ② B10 の式を B11:B13 に複写する。

- ③ C10 に次のような式を入力する (但し 売上!\$E\$12:\$U\$18 は規模分類の値が 2 であるレコードのすべての商品コードを含む範囲)。

$$=countif(売上!E12:U18, 商品分類別!A10 & "*")$$

- ④ C10 の式を C11:C13 に複写する。

- ⑤ D10 に次のような式を入力する (但し 売上!\$E\$19:\$U\$26 は規模分類の値が 3 であるレコードのすべての商品コードを含む範囲)。

=countif(売上!\$E\$19:\$U\$26,商品分類!A10 & "*"")

⑥ D10の式をD11:D13に複写する。

〈例5.10.2〉規模分類と商品分類の組み合わせごとに売上合計を求めるには

① B16に次のような式を入力する(但し 売上!\$H\$9:\$X\$11は規模分類の値が1であるレコードのすべての小計を含む範囲)。

=sumif(売上!\$E\$9:\$U\$11,商品分類!A16 & "*",売上!\$H\$9:\$X\$11)

② B16の式をB17:B19に複写する。

③ C16に次のような式を入力する(但し 売上!\$H\$12:\$X\$18は規模分類の値が2であるレコードのすべての小計を含む範囲)。

=sumif(売上!\$E\$12:\$U\$18,商品分類!A16 & "*",売上!\$H\$12:\$X\$18)

④ C16の式をC17:C19に複写する。

⑤ D16に次のような式を入力する(但し 売上!\$H\$19:\$X\$26は規模分類の値が3であるレコードのすべての小計を含む範囲)。

=sumif(売上!\$E\$19:\$U\$26,商品分類!A16 & "*",売上!\$H\$19:\$X\$26)

⑥ D16の式をD17:D19に複写する。

例5.10.1と例5.10.2を実行すると、図5.17のような結果が得られる。

	A	B	C	D
9	件数	1	2	3
10	A	3	3	2
11	B	1	2	8
12	C	6	4	5
13	D	0	5	1
14				
15	売上	1	2	3
16	A	85000	22000	13950
17	B	16500	23700	10650
18	C	31500	17800	9200
19	D	0	465000	7500

(図5.17) 例5.10.1と例5.10.2の実行結果

しかし、例5.10.1と例5.10.2の操作は、実際に操作してみるとわかるように、かなり煩雑で操作ミス誘発しやすい。この場合も、もし図5.12のような第1正規形データがあるならば、それにピボットテーブルを適用した方がはるかに楽に集計をすることができる。

6. おわりに

以上の考察をまとめると、次のようなことが言えよう。

- 集計フィールドが非繰り返しフィールドのときは、場合により補助フィールドを付加するな

どして非正規形データに直接ピボットテーブルを適用することができる。

- 集計フィールドが繰り返しフィールドのときは、countif 関数や sumif 関数を駆使して非正規形データから集計できる場合もある。しかし、それと同等の第 1 正規形データがあれば、それに直接ピボットテーブルを適用する方が簡単である。

したがって、非正規形データが与えられているときは、それを基に第 1 正規形データを生成して、両者を併用するのが現実的であろう。しかし、データが大量にある場合は、そのような変換を手動で行うのは困難であるので、Excel VBA などによる変換プログラムが必要になるであろう。そのような変換プログラムや、RDBMS による集計作業との比較については、別の機会に論じてみたい。

参考文献

- [1] 芝野耕司『SQL がわかる本』オーム社，1998.
- [2] 鈴木昭男『Access95 で学ぶリレーショナル・データベース』ソフト・リサーチ・センター，1996.
- [3] 田中ナルミ・阿部忠光『標準 MySQL』ソフトバンク，2003.