





統計学で使う表計算ソフトの解説

1 表計算ソフトの概念

よく用いられる表計算ソフトには、マイクロソフト (Microsoft) 社のエクセル (excel)  や、サン (Sun Microsystems) 社のオープンオフィス=カルク (open-office calc)  MySQL   があります。前者は購入する必要がありますが、後者はフリーウェアですから、規約を守ることですべて使えます。

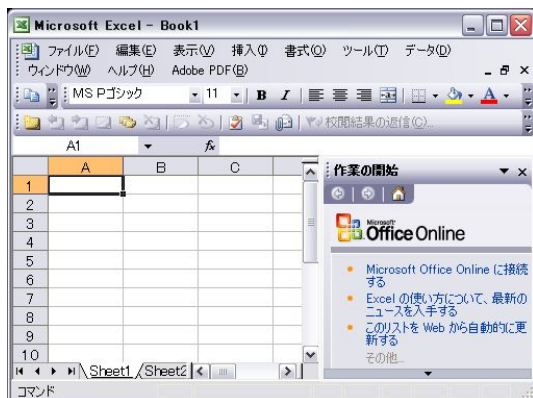


図 1 エクセルの立ち上がり画面

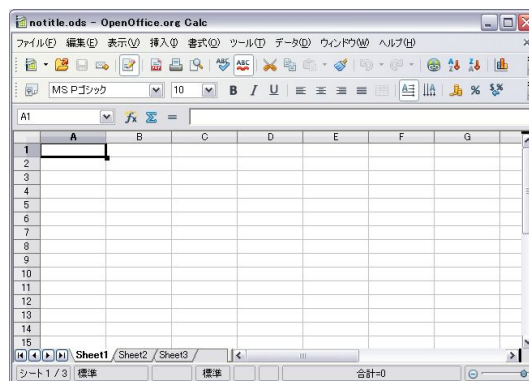
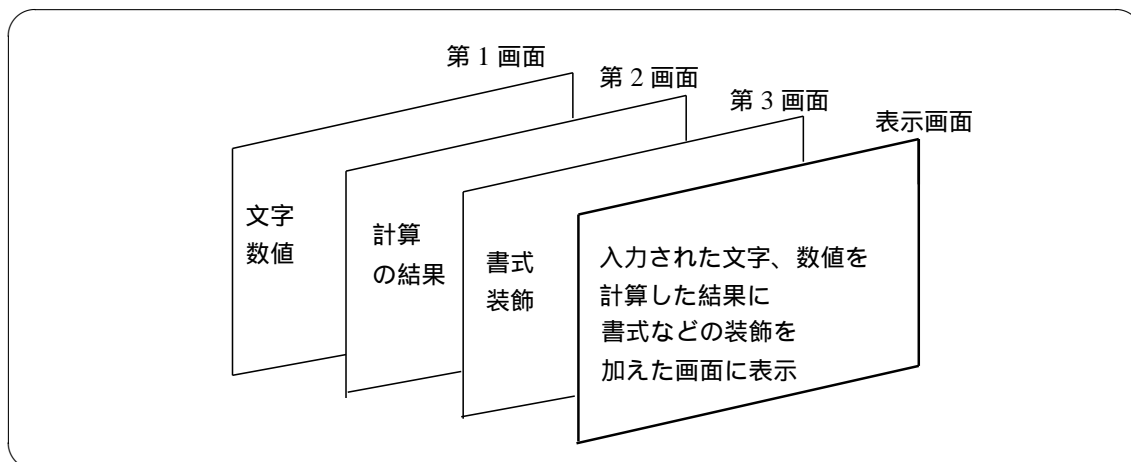


図 2 オープンオフィスのカルク

表計算ソフトの画面には、3つの階層がまとめられて表示されると考えれば理解し易いでしょう。その3つの層とは、つぎのものです。表計算ソフトの1枚のシートには、3層と表示領域から構成される。多くの機能を使いこなすことは、データの整理やプレゼンテーションにはたいへん役に立ちます。

- 第1層には文字、数値、数式が値の規則として入力
- 第2層には数式による計算結果が値として入力
- 第3層には表示形式、書式など、さらに表示文字色などで修飾した結果が入力
- 最終の表示領域にはこれらの結果が表示される



以下では、統計ではよく用いられるデータ集計の整理のために、度数分布表とそのグラフ化 (ヒストグラム) をつくる方法の説明を述べます。

2 度数分布表-countif

Excel のグラフを用いてヒストグラムを作成する方法を紹介します。

ヒストグラムとは、データの度数分布をグラフにしたものです。データがどのような分布をしているかを知るためのもっとも簡便でわかりやすいグラフです。これから紹介する方法を用いて 22 図のようなヒストグラムを作成します。

参考：統計WEB <http://software.ssri.co.jp/statweb2/> - アマチュアのための統計学講座 -

集計データの作成

1. ある統計学試験を 111 人に行った結果として得られたデータ結果：図 3 から、得点の階級別の集計とそのヒストグラムを作成します。
2. 図 4 のように度数分布表の枠組みをまず準備作成します。

	A	B	C	D
1				
2		統計学試験の得点		
3		得点		
4		87		
5		87		
6		31		
7		20		
8		66		
9		41		
10		100		
11		67		
12		92		
13		96		
14		00		

図 3 入力されたデータ

	A	B	C	D	E	F	G
1							
2		統計学試験の得点					
3		得点	階級値	階級名	度数		
4		87	0				
5		87	1				
6		31	20				
7		20	40				
8		66	60				
9		41	80				
10		100	100				
11		67		合計			
12		92					

図 4 度数分布表の枠組み

3. 「階級名」のセル範囲「E4:E10」の書式設定を行います。セル範囲「E4:E10」を選択後、右クリックメニューから「セルの書式設定」ウィンドウを開きます。[表示形式] タブで「分類」を[文字列]とし、[OK]をクリックします。文字列として設定している理由はセルの中に用いるハイフン「-」を数式のマイナス「-」とは異なった意味で用いるためです。
4. 「階級名」のセル範囲「E4:E10」の書式設定を行います。セル範囲「E4:E10」を選択後、右クリックメニューから「セルの書式設定」ウィンドウを開きます。[表示形式] タブで「分類」を[文字

	A	B	C	D	E	F
1						
2		統計学試験の得点				
3		得点	階級値	階級名	度数	
4		87	0			
5		87	1			
6		31	20			
7		20	40			
8		66	60			
9		41	80			
10		100	100			
11		67		合計		
12		92				

図 5 階級を「文字」として設定

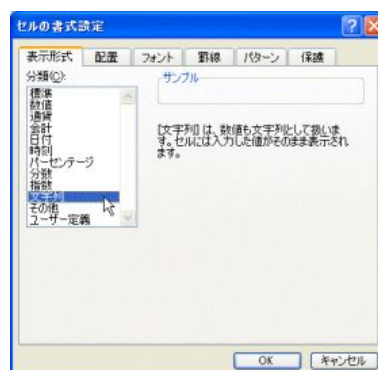


図 6 書式設定で文字列を選択

列]とし、[OK]をクリックします。文字列として設定している理由はセルの中に用いるハイフン「-」を数式のマイナス「-」とは異なった意味で用いるためです。

5. 階級名を入力します。前の手順で表示形式を設定したので、「1-19」のような階級名が日付や計算などに自動変換されることはありません。
6. 各階級の度数を Excel の COUNTIF 関数を用います。COUNTIF 関数の引数は、「データ範囲」と「データを数えるための条件」です。「100」に等しいというデータの件数を数える場合、=COUNTIF(データ範囲,100)とします。「20以上」のデータの件数を数える場合、=COUNTIF(データ範囲,">= 20")とします。「20以上40未満」のデータの件数を数える場合、=COUNTIF(データ範囲,">= 20") - COUNTIF(データ範囲,">= 40")とします。このように論理式が成り立つものの個数を計算します。なお、度数の合計を「合計」の欄に計算し、元データの件数と同じ値であることを確認しておきます。この計算方法は、ある区間における度数の計算は2つの裾部分での差で実現しています。

	A	B	C	D	E	F	G
1							
2		統計学試験の得点					
3		得点	階級値	階級名	度数		
4		87	0	0			
5		87	1	1-19			
6		31	20	20-39			
7		20	40	40-59			
8		66	60	60-79			
9		41	80	80-99			
10		100	100	100			
11		67		合計			
12		92					

図7 階級名が入りました

	A	B	C	D	E	F	G
1							
2		統計学試験の得点					
3		得点	階級値	階級名	度数		
4		87	0	0	1		
5		87	1	1-19	22		
6		31	20	20-39	13		
7		20	40	40-59	24		
8		66	60	60-79	28		
9		41	80	80-99	21		
10		100	100	100	2		
11		67		合計	111		
12		92					

図8 度数の集計結果

つぎに縦棒グラフとして作成

7. セル範囲「E3:F10」を選択後、「グラフウィザード」ボタンをクリックします。左側の「グラフの種類」の欄で[縦棒]を選択し、右側の「形式」の欄で[集合縦棒]を選択して、[次へ]をクリックします。(9図参照)
8. [グラフの元データ]ウィンドウではこのまま[次へ]をクリックします。(10図参照)
9. [タイトルとラベル]タブで「グラフタイトル」を入力します。(図参照)



図9 まず棒グラフを選択



図10 ウィザードの段階



図 11 グラフオプション (凡例あり)



図 12 オプション「凡例なし」に設定

10. [凡例] タブで [凡例を表示する] のチェックを外し、[完了] をクリックします。(図参照)
この時点でグラフは図のようになります。

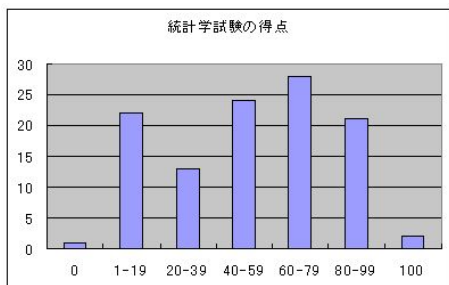


図 13 一時的なヒストグラム



図 14 棒の間隔をゼロに設定

棒の間隔を調整する

ここからは「グラフ」ツールバーを用いてグラフの設定を行うと便利です。グラフを選択しても「グラフ」ツールバーが表示されない場合、グラフを選択した状態で、メニューより [表示] [ツールバー] [グラフ] を選択してください。

11. 系列「度数」の書式設定ウィンドウを開き、[オプション] タブで「棒の間隔」に「0」を入力します。間隔調整する理由はヒストグラムの幅に意味をもたせるためです。
オープンオフィスでは、棒に合わせて右クリック、オブジェクトの属性に 「データ系列」「オプション」「設定」「間隔 (S)」「0 %」で OK をクリック

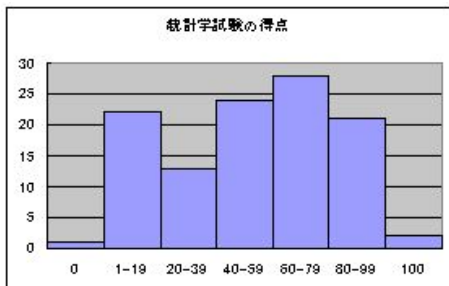


図 15 ヒストグラムの完成

A	B	C	D	E	F	G	H	I
1								
2		統計学試験の得点						
3	得点	階級値	階級名	度数	相対度数	累積相対度数		
4	87	0.0		1	0.009	0.009		
5	87	1.1-1.9		22	0.198	0.207		
6	31	2.0-2.9		13	0.117	0.324		
7	20	4.0-4.9		24	0.216	0.541		
8	66	6.0-6.9		28	0.252	0.793		
9	41	8.0-8.9		21	0.189	0.982		
10	100	10.0		2	0.018	1.000		
11	67			合計		111		
12	92							

図 16 累積度数の計算

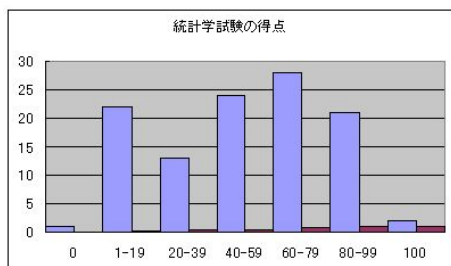


図 17 最初にグラフを貼り付け



図 18 グラフを追加すると

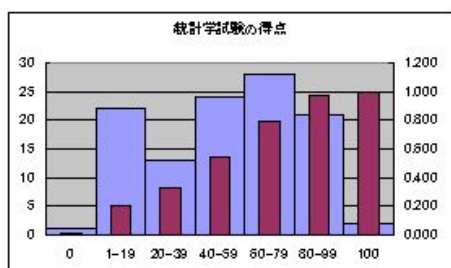


図 19 表示の方法を変えていく



図 20 折れ線グラフに変更

累積相対度数を追加する方法

12. 度数分布表の右側に相対度数と累積相対度数を求めます。相対度数は各階級の度数を合計で割った値、累積相対度数は相対度数を最初の階級から順に足し上げていった値です。対応する値より小さい部分が全体のうちのどの位を占めているかを表します。
13. 累積相対度数のセル範囲「H3:H10」をコピーします。グラフの余白で右クリックし、[貼り付け]を選択します。
14. 系列「累積相対度数」の書式設定ウィンドウを開きます。[軸]タブで「表示する軸」として[第2軸]を選択し、[OK]をクリックします。
15. 累積相対度数の棒グラフの上で右クリックし、[グラフの種類]を選択します。左側の「グラフの種類」の欄で[折れ線]を選択し、右側の「形式」の欄で「データにマーカーが付けられた折れ線グラフ」を選択して、[OK]をクリックします。
ヒストグラムと相対度数曲線（多角形）の完成
16. 適宜グラフの書式設定を行い、ヒストグラムの完成です。

最後にヒストグラムは階級値の値を変えるとグラフの形が変わります。階級の区間幅、個数は最大値、最小値、データの数から考えます。データ数と階級の個数の全体像をバランスよく把握することは探索的なデータ解析の基礎で、試行錯誤しながら、またまたスタージェスの公式 $k = 1 + \log_2 n$ なども

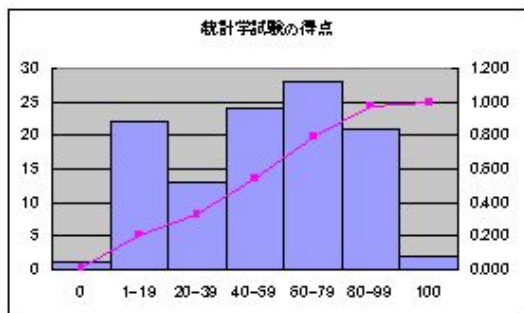


図 21 ヒストグラムと累積相対度数曲線

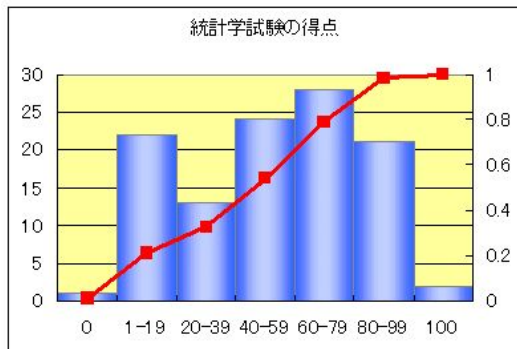


図 22 オプションで見栄えよく

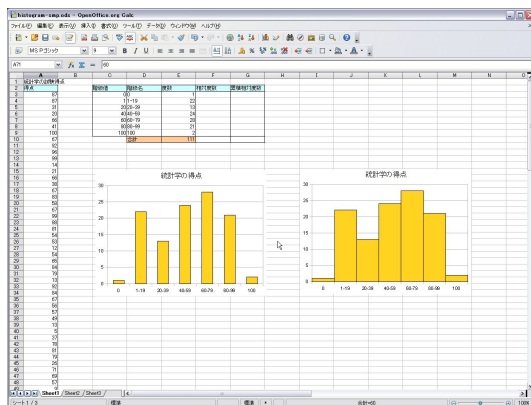


図 23 オフィスカルクでの操作 (間隔を変更)

考えられています。例えば、観測値の数 n を 100 とすると、スタージェスの公式では $k = 5.6$ となります。多くの場合、階級の個数 $k = 6$ くらいが見た目がよいという目安です。今回は省いていますが、本来はもっとも適切な階級値の幅とその個数を探る必要があるでしょう。

3 正の字で集計

例題：観測データを整理する方法ではつぎの表も使われます。

項目	正マーク	tally mark	値
A	T		2
B	正正下		13
C	正正		9
D	正正正正T		22
E	下		3
F	一		1
計		50	

4 度数分布表-pivot table

ピボットテーブルで度数分布表あるいはヒストグラムをつくります。度数分布表は離散型の変数値に対する頻度 (比率) を対応させ、ヒストグラムは連続型の変数値に対応させます。データのとる値は多く

ないときには、そのままの形でも大まかな形状は理解できますが、変域が広い場合などには、階級に分類させて整理したほうが分かり易くなります。

ピボットテーブルでは「グループ化」のやり方、またグラフの表示では、とりあえず棒グラフを選択してから、値の間隔をつめていき、ヒストグラムにつくります。

【例題】

列 \ 行	A
1	総得点
2	150
3	200
4	250
5	320
6	330
7	360
8	380
9	420
10	480

セル範囲 "A2:A10" には左表のように総得点が入力されているとします。このデータについて、右の表に (1) 行から (4) 行がそれぞれ何件あるかをピボットテーブルで集計するにはどうすれば良いでしょうか？

	総得点の階級分け	度数 (比率)
(1)	200 未満	?
(2)	200 以上 300 未満	?
(3)	300 以上 400 未満	?
(4)	400 以上 500 未満	?
	計	?

【解】

1. リスト内 (A1:A10) のセルのどれかを選択
2. メニュー [データ]-[ピボットテーブルとピボットグラフレポート]
3. [Excel のリスト / データベース] と [ピボットテーブル] にチェックが入っていることを確認して、[次へ] ボタンをクリック
4. [範囲] ボックスに \$A\$1:\$A\$10 が入力されていることを確認して [次へ] ボタンをクリック、そして [レイアウト] ボタンをクリック
5. [行] に 総得点 をドラッグ、[データ] にも 総得点 をドラッグ
6. [データ] にドラッグしたデータの個数 : 総得点 をダブルクリック
7. [集計の方法] で データの個数 をダブルクリック、[OK] ボタンをクリック
8. [既存のワークシート] をクリック。例えばセル D1 をクリック。
9. [完了] ボタンをクリック → これで、いったんピボットテーブルが完成、つぎにセル範囲の表示をグループ化する
10. セル D2 上で右クリック-[グループ化]
11. [先頭の値] を 200 に、[末尾の値] を 500 にまた [単位] が 100 を確認
12. [OK] ボタンをクリック、これで のように表示されます。

5 度数分布表 - frequency

組み込み関数 frequency 関数をもちいる方法を説明します。ウィザードに従えばできますが、最後の OK をクリックするとき、数式配列の入力に注意します。

1. 予めデータの最大値 (max) と最小値 (min) をもとめて、データの個数から階級の幅と各階級の始まりと終わりを求める。
2. これを D13 から D17 までに入力 (連続データの作成)

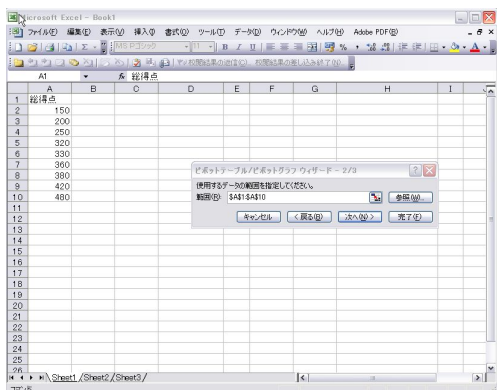


図 24 グラフを作るウィザード 01

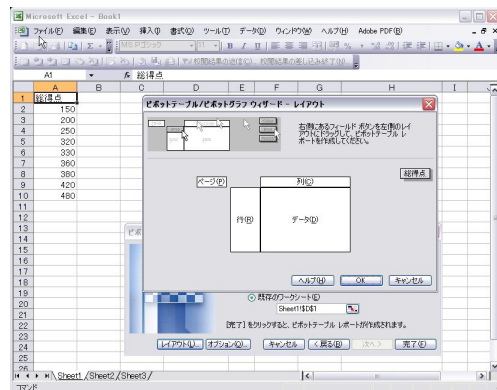


図 25 グラフを作るウィザード 02

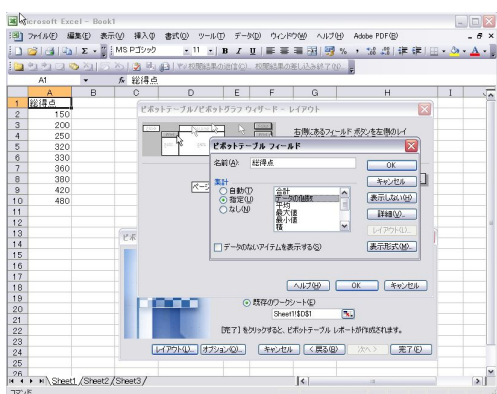


図 26 グラフを作るウィザード 03

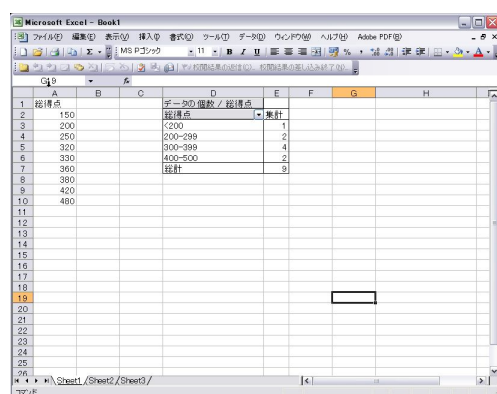


図 27 グラフを作るウィザード 04

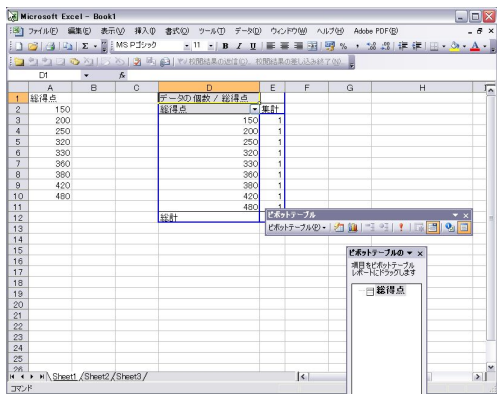


図 28 グラフを作るウィザード 021

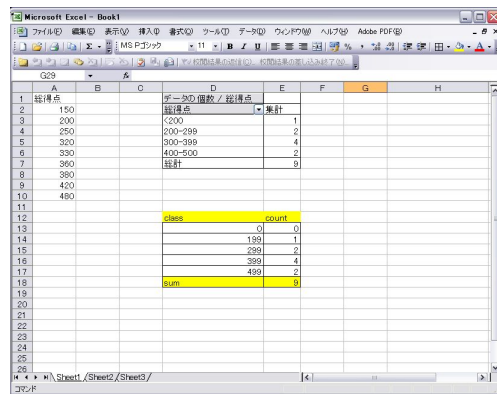


図 29 グラフを作るウィザード 04

3. 度数 (集計) を記入する E13 から E17 までのセルをアクティブに
4. 数式バーに=frequency と書き
5. 組み込み関数 FREQUENCY(データ配列, 区間配列) については、データ配列が A2 から A10 とし、区間配列には D13 から D17 までとする
6. ここで単に OK をクリックするのではなく、数式配列の入力 CTRL+SHITFT を押しながら OK をクリックする。

いずれの方法にしても答えは次の通りです。

データの個数 / 総得点				
	総得点の階級分け	集計	度数	比率
(1)	<200	—	1	0.11
(2)	200-299	T	2	0.22
(3)	300-399	正	4	0.44
(4)	400-500	T	2	0.22
	計		9	1.00

ここで度数から比率の計算をするためには、度数の合計を”=sum(A2:A9)”で求め、このセル値を「絶対参照」してそれぞれの度数を割っていきます。これは各度数が「相対参照」され、分母が絶対参照ですから、コピーとペーストでドラックすれば計算できます。

【演習のための課題】2変量データの場合について、組み込み関数 countif と論理関数 AND を組み合わせると、2次元の度数分布表が求められます。これを考えてみて下さい。

カーソル, セルの移動	矢印キー, Ctrl, Shift の組み合わせ	shift+ ctrl+ arrow key
セルの選択	active, nonactive	
データの並び替え		
データの抽出		

表 1 命令

つぎの2つ図は、「エクセル」と「オープンオフィスのカルク」でのグラフを作るそれぞれのウィザードを示しています。ほぼ同じようにして作成することができます。



図 30 エクセルのグラフを作るウィザード



図 31 オープンオフィスのカルクでグラフを作る

グラフ化	範囲の選択 グラフの選択 パラメータオプション
棒グラフ	一定間隔の数値の変化や傾向を調べ、項目間を比較する。加算に意味をもつデータに適用
折れ線グラフ	データの変化の状況を表し、いくつかの傾向を同時に表示できる。とくに時間変化の傾向を調べたりする。
円グラフ	それぞれの項目間の値に対して、全体に対する割合を比較する。面積比で数値の比較
散布図 (相関図)	2 変量を縦軸と横軸にして、収集したデータをカテゴリー化しないで座標軸にプロットする。もしデータがいくつかの群に分けられているときには、グラフ内の点の記号を変えてプロットする。複数の変数観の散布図をまとめて表示するれば、相関の強弱が視覚的に捉えられる。
レーダーチャート	複数のデータ系列の構成成分を比較表示する。各系列間のバランスや描かれた多角形の面積の数値により、多次元 (多変量) データ間の比較ができる
その他のグラフ	層グラフ (連続するデータの合計と内訳の推移を表す)

表 2 グラフのいろいろ

表計算ソフトの命令		例
四則演算記号 (加減乗除)	+ , - , * , /	
比較演算記号 (大小、等値)	= , < , > , < = , > = ,	
セル番地	セルの行 (数字) と列 (英字) で指定	セルは列をアルファベットの並び順, 行を番号付けした座標形式で表す
セル範囲 (矩形領域の表現)	開始セル番地 ~ 終了セル番地	
絶対参照	列あるいは行を表す英数字の前に記号 \$ をつけて使う	相対参照は数式を作成するセルを基点として座標を指定する
複写 (コピー)	セルやセル範囲の参照を含む計算式を複写した場合, 相対的な位置関係が保たれるように参照する列、行が変更される	
ワークシート参照	別のワークシートでのセルやセル範囲を参照するには, ワークシート名と記号 ! をつける	=関数名 (引数 1, 引数 2,, 引数 n)
数式 (関数の構成)	等号で始めて, セル間の演算や組み込み関数を用いた式	
式 (論理式, 条件式)	数値、セル間の関係や数式の命題、条件を表すもの	
I F (条件式, 式 1, 式 2)	条件式が成り立つ (真) とき, 式 1 をの値を返し, そうでないとき (偽) には式 2 の値を返す	条件式は論理式ともいいかえられる
A N D (条件式 1, 条件式 2,, 条件 n)	条件式 1, 条件式 2,, 条件 n がすべて成り立つとき, 真を返し, そうでないとき偽を返す	
O R (条件式 1, 条件式 2,, 条件 n)	条件式 1, 条件式 2,, 条件 n のうち一つでも成り立つとき, 真を返し, そうでないとき偽を返す	
組み込み関数	あらかじめプログラムで登録定義されている関数	関数を利用するには, 関数名の前に等号 (=) をつけるが, 引数として使うときには等号はつけない
真 (true)	条件の命題が満たされる, あるいは成り立つとき。	
偽 (false)	条件の命題が満たされない, あるいは成り立たないとき	
等号 (=)	条件命題では等しい、同じことを意味する。さらに値や式の結果値を代入する、割り当てる場合にももちいる。	

表 3 命令の説明

より高度な関数	形式	用例など
I F 関数	= I F (論理式 , 値 1 , 値 2)	論理式が真の場合は値 1 を偽の場合には値 2 を返す
A N D 関数	= A N D (論理式 1 , 論理式 2 , ...)	= I F (A N D (論理式 1 , 論理式 2 , ...) , 真の場合 , 偽の場合)
O R 関数	= O R (論理式 1 , 論理式 2 , ...)	= I F (O R (論理式 1 , 論理式 2 , ...) , 真の場合 , 偽の場合)
S U M 関数	= S U M (引数 1 , 引数 2 ,)	引数で定められる数値や範囲について , 数値の合計を返す
S U M I F 関数	= S U M I F (範囲 , 検索条件 , 合計範囲)	範囲 : 検索条件で評価される対象となるセル 範囲検索条件 : 計算の対象となるセルを定義する条件 合計範囲 : 実際に計算の対象となるセルの範囲
C O U N T 関数	= C O U N T (値 1 , 値 2 ,)	
C O U N T I F 関数	= C O U N T I F (範囲 , 検索条件)	

表 4 命令