

I Excel を用いたデータ解析と情報教育

土屋 高宏*

TSUCHIYA Takahiro

1. はじめに

Microsoft 社の Excel は 1970 年代後半、当時ハーバード大学のビジネススクールで学んでいたダン・ブルックリンによってその基本的なアイデアが創案され、表計算ソフトとしてこの世に生まれました。Excel は表計算からグラフ作成、データ解析に至るまでさまざまな機能を兼ね備えており、最新では Excel 2003 が発売されています。Microsoft Office 製品は世界的にもメジャーであり、Excel の基礎や応用を解説したテキストは数多く出版されています。ここでは、Excel の基本操作から始めて、データの処理・解析の基本的な手法を実習問題を通して紹介するとともに、Excel の中学・高校数学教育への利用、さらには教育現場で必要となる成績処理などへの適用について検討します。なお Excel のバージョンは 2004 年 8 月現在、城西大学情報科学研究センターの情報教育システム(SCNL)に導入されている Office 2000 Professional を想定しています。

2. Excel の起動と終了

SCNL のシステムでは、[スタート]→[プログラム(P)]→[Microsoft Office]→[Microsoft Excel]を選択すると Excel が起動します。Excel を終了するには、メニューの[ファイル]→[終了]を選択するか、ウィンドウの右上端にある[閉じる]ボタンをクリックします。

3. Excel の基本操作とデータの入力

表計算やグラフ、データベースの作成や操作を行う領域を**ワークシート**といいます。1つのワークシートは 65,536 行×256 列から構成され、その一部がウィンドウとして表示されます。ワークシート上にある 1つ1つのます目を**セル**といい、1つのセルには1つのデータを入力することができます。セルの位置は列(アルファベット)と行(数字)との組合せで表されます。例えば、セル A 1 は

* 城西大学理学部数学科

A列の1行目にあるセルを表します。入力を行うことができるセルを**アクティブセル**とよび、まず目が太い線で囲まれて表示されます。アクティブセルはキーボードの4つの矢印(←→↑↓)を操作するかマウスを使って目的のセルにマウスポインタを移動させ、クリックすることによって任意のセルに移動させることができます。また、ウィンドウに現れていないセルに移動するには、スクロールバーかマウスのスクロールボタンを使うと便利です。

Excel では数値や文字列、計算式、関数などのデータを入力することができます。セルに文字や数値を入力したら、内容を確定させるために **Enter** キーを押します。

実習1 セルA1からA5に次のデータを入力してください。

(1) Josai (2) じょうさい (3) ジョウサイ (4) 城西 (5) 城西大学理学部

3. 1 セルの幅の変更

列番号の右側の境界線にマウスポインタを移動させ、 \leftrightarrow に変わったところで必要な幅になるまでドラッグ(マウスの左ボタンを押しながら移動)します。また、列番号の右側の境界線 \leftrightarrow を**ダブルクリック**すると、列内の最長のデータに合わせて、列幅を自動的に調整できます。

3. 2 セルの内容の修正と消去

すでに入力されているセルの内容を修正するには、マウスポインタを内容を修正したいセルへ移動させ、ダブルクリックするか **F2** キーを押してセルを編集状態にします。また、セルの内容を消去するには **Delete** キーを押します。

実習2 セルA4の内容を**城西大学**と修正し、A3の内容を消去してください。

3. 3 計算式の入力

Excel では、電卓の代わりにいろいろな計算を行うことができます。例えば、 $(2.4+5.3) \times 3 - 8.6$ や $(1+3)^2 \div 5$ はセルに

$$=(2.4+5.3)*3-8.6$$

$$=(1+3)^2/5$$

と入力します。計算の優先順位は、通常の数学記号と同様「べき乗」→「掛け算、割り算」→「足し算、引き算」で、優先順位を変えたい場合は () を使います。{ } や [] は使わず、複数個のカッコが必要な場合は () を複数個使います。また、Excel には多くの算術関数が用意されています。関数名が分からないときなどには、メニューの[ヘルプ(H)]→[Microsoft Excel ヘルプ(H)]を選択して調べることができます。

表1 算術演算子と算術関数

入力	意味	入力	意味	入力	意味	入力	意味
+	加算	=EXP(x)	e^x	=TAN(x)	$\tan x$	=ATAN(x)	$\arctan x$
-	減算	=LN(x)	$\log x$	=SQRT(x)	\sqrt{x}	=PI()	π
*	乗算	=LOG10(x)	$\log_{10} x$	=FACT(n)	$n!$	=SINH(x)	$\sinh x$
/	除算	=SIN(x)	$\sin x$	=ASIN(x)	$\arcsin x$	=COSH(x)	$\cosh x$
=X^n	x^n	=COS(x)	$\cos x$	=ACOS(x)	$\arccos x$	=TANH(x)	$\tanh x$

実習3 次の計算してください。

- (1) $5 + 16 - 7$ (2) $(3 + 9 \times 5) \div 6$ (3) $3 \div 4^2$ (4) $3\{2 + (3 + 2)^3\}$
 (5) $e^{1.7}$ (6) $\sqrt{5 - \sqrt{24}}$ (7) $2 \times \log_{10} 5$ (8) $\sin^2 0.5 + \cos^2 0.5$
 (9) $\tan \frac{\pi}{3}$ (10) $\left| \log \frac{1}{2} \right|$ (絶対値を求める関数は=ABS(x))

3.4 ワークシートの保存

作成したワークシートをファイルとして保存するには、メニューの[ファイル(F)]→[名前を付けて保存(A)]をクリックしてください。「ファイル名を付けて保存」ダイアログボックスが表示されますので、保存先を選択し、ファイル名を入力します。

3.5 連続データの入力

数値や日付などの連続データを入力するには、**オートフィル機能**を用いると便利です。例えば、1月、2月、3月を入力する場合、まず最初の月の1月（初期値）を入力します。1月と入力したセルの右下隅にマウスポインタを合わせると「+」に変わりますので、3月になるまで入力するセル範囲をドラッグして指定すると、1月から3月までの月が自動的に埋め込まれます。「月」以外にも「年」、「日」、「曜日」、「干支」、「第1期、第2期、…」などのデータも、オートフィル機能を利用して簡単に入力することができます。また、1、2、3、…、10を入力する場合、最初の2つの数字の1と2を初期値として入力し、2つの初期値が入力されているセル範囲をドラッグして指定した状態でセルの右下隅にマウスポインタを合わせ、10になるまでドラッグすると、1から10までの数字が入力されます。

3.6 罫線を引き方

表の範囲をドラッグして指定し、メニューの[書式(O)]→[セル(E)]をクリックすると、「セルの書式設定」ボックスが表示されますので、[罫線]を選択して目的に応じた罫線を引くことができま

す。また、書式設定ツールバーの[罫線]ボタンから簡単な罫線を引くこともできます。

3. 7 行や列の挿入・削除

行番号または列番号を右クリックするとショートカットメニューが表示されますので、[挿入(I)]を選択すると行または列が挿入され、[削除(D)]を選択すると行または列が削除されます。

4 Excel による表計算

Excel の最も重要な機能の1つは表計算機能です。式や関数をうまく使うことによって、計算や集計を容易に行うことができます。例えば、セルA 1からC 5に次のようなデータを入力し、行の合計を計算してみましょう。

1	6	11
2	7	12
3	8	13
4	9	14
5	10	15

合計の計算には次の3つの方法があります。例として、セルD 1にA 1, B 1, C 1の合計を計算してみます。

① 計算式を入力する。D 1に=A1+B1+C1 と入力し、 キーを押します。

② SUM 関数を用いる。D 1に=SUM(A1:C1)と入力し、 キーを押します。あるいは、標準ツールバーの[関数貼り付け]ボタン をクリックすると、「関数の貼り付け」ボックスが表示されますので、[数学/三角]→[SUM]を選択し、[OK]をクリックします。[数値1]にデータ範囲A1:C1を指定します。データ範囲はセルA 1をクリックし、セルC 1までドラッグすることにより指定することができます。

③ オートSUMボタンを用いる。D 1をクリックし、標準ツールバーの[オートSUM]ボタン をクリックします。セルA 1をクリックし、C 1までドラッグします。最後に キーを押します。

同様にD 2からD 5まで行の合計を計算してみます。=A1+B1+C1 や =SUM(A1:C1)などと同じような入力を繰り返していくのでは面倒ですし、間違いの原因となります。この場合、オートフィールド機能を使うと入力を繰り返すことなく合計の計算を行うことができます。アクティブセルをすでに合計が入力されているセルD 1へ移動させ、セルの右下隅にマウスポインタを合わせると「+」に変わります。セルD 5までドラッグするとD 1の内容がD 2からD 5までにコピーされ、すべての行の合計が計算されます。

参考 「+」をダブルクリックすると隣接する列の最終行を自動的に認識し、数式のコピーが容易にできます。縦に入力する範囲が長く、ドラッグしづらい場合に使うと便利です。

実習 4 A 6 に A 1 から A 5 までの合計を計算し、B 列と C 列の合計をオートフィル機能を使って計算してください。また、E 1 に A 1 から C 1 までの平均 $=D1/3$ を計算し、これをコピーして E 2 から E 5 に各行の平均を計算してください。

平均は AVERAGE 関数を使って求めることもできます。A 1 から C 1 までの平均は $=AVERAGE(A1:C1)$ により計算することができます。Excel には平均を求める AVERAGE のように、データの集計や統計処理に使うことができる関数が数多く用意されています。これは、標準ツールバーの[関数貼り付け]ボタン  をクリックし、「関数の貼り付け」ボックスの[統計]から選択することができます。主な統計関数は表 2 の通りです。

表 2 主な統計関数

表記	意味
AVERAGE(範囲)	範囲のデータから平均を求める
CORREL(範囲 1, 範囲 2)	範囲 1 と範囲 2 のデータから相関係数を求める
COUNT(範囲)	数値の入ったセルの数を求める
COUNTA(範囲)	空白でないセルの数を求める
COUNTIF(範囲, 検索条件)	範囲のセルのうち検索条件のセルの数を求める
COVAR(範囲 1, 範囲 2)	偏差積和をデータの数で割った共分散を求める
DEVSQ(範囲)	偏差平方和を求める
MAX(範囲)	最大値を求める
MEDIAN(範囲)	中央値を求める
MIN(範囲)	最小値を求める
QUARTILE(範囲, 数値)	数値で指定された四分位点を求める
STDEV(範囲)	偏差平方和をデータの数 - 1 で割った標準偏差を求める
STDEVP(範囲)	偏差平方和をデータの数で割った標準偏差を求める
VAR(範囲)	偏差平方和をデータの数 - 1 で割った分散を求める
VARP(範囲)	偏差平方和をデータの数で割った分散を求める
SKEW(範囲)	データの歪度(データの分布の偏りに関する指標)を求める
KURT(範囲)	データの尖度(データの分布の尖りに関する指標)を求める

実習5 次の表を完成させ、各観測地点の4月から7月までの合計降水量と平均降水量を計算し、合計降水量は小数点以下1桁まで、平均降水量は小数点以下2桁まで表示してください。また、統計関数を用いて各月の最大降水量と最小降水量を求めてください。(小数点表示は表示させたいデータ範囲をドラッグして指定し、書式設定ツールバーの[小数点表示桁上げ]ボタン、あるいは[小数点表示桁下げ]ボタンを uses。)

月別平均降水量(mm)

観測地点	4月	5月	6月	7月	合計降水量	平均降水量
札幌	62.3	54.8	66.4	68.7		
仙台	93.8	108.8	133.1	150.7		
東京	125.0	138.0	185.2	126.1		
名古屋	150.5	157.1	217.9	212.5		
大阪	133.8	139.4	206.4	156.9		
広島	139.6	155.8	232.4	258.4		
福岡	126.5	144.1	256.2	257.6		
最大降水量						
最小降水量						

【理科年表平成13年国立天文台編より】

5 1次元データの解析

次のデータは生徒50人の数学の試験の得点を表しています。

76	82	48	68	92	77	32	43	78	96
85	61	70	88	57	73	56	76	66	75
84	80	51	92	80	38	69	71	65	45
67	81	98	63	74	60	57	78	80	54
71	78	84	65	52	68	74	83	56	66

このデータを整理してみましょう。1次元データの整理・分析には次の2つの方法があります。

1. 度数分布表やヒストグラムを作成する。
2. データのもつ特性を数量的に要約する。

5.1 度数分布表とヒストグラム

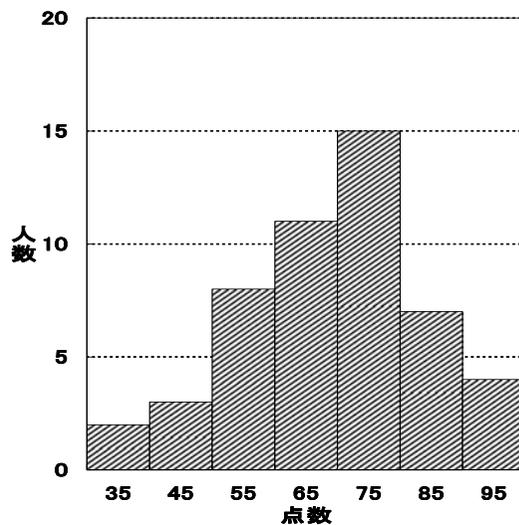
度数分布表は観測された n 個のデータのとりうる値をいくつかの範囲(**階級**)に分け、その階級に属するデータの個数(**度数**)を数えて表にしたものです。 k 個の階級に分けた場合の度数分布表は次のようになります。

階級 (より大きい)～(以下)	階級値 m_i	度数 f_i	相対度数 f_i/n
$a_0 \sim a_1$	m_1	f_1	f_1/n
$a_1 \sim a_2$	m_2	f_2	f_2/n
\vdots	\vdots	\vdots	\vdots
$a_{k-1} \sim a_k$	m_k	f_k	f_k/n
計		n	1

階級値は各階級の中央の値で、**相対度数**は各階級の度数を全体のデータ数で割ったものです。**ヒストグラム**は度数分布表を図で表現したもので、各階級を横軸にとり、階級の幅を底辺とし度数を高さとする長方形を描くことによって作ることができます。上記データを7個の階級に分類して度数分布表とヒストグラムを作成すると、以下のようになります。

階級	階級値	度数	相対度数
30～40	35	2	0.04
40～50	45	3	0.06
50～60	55	8	0.16
60～70	65	11	0.22
70～80	75	15	0.30
80～90	85	7	0.14
90～100	95	4	0.08
計		50	1.00

試験得点のヒストグラム



実習 6 以下の手順にしたがって、成績データの度数分布表とヒストグラムを作成してください。

はじめに[分析ツール(D)]が組み込まれているかどうかを確認しましょう。メニューの[ツール(T)]に[分析ツール(D)]がない場合は[アドイン(I)]をクリックします。「アドイン」ボックスが表示されますので、[分析ツール]の前のボックスをチェックし[OK]をクリックすると、分析ツールが組み込まれます。

セルA 1に**数学の成績**、A 2とB 2にデータの項目名として**生徒番号**、**得点**と入力してください。

① 第3行から第52行にデータを入力します。

このデータの生徒の数は50人ですので、1から50までの数字を**生徒番号**と入力したセルの1つ下のA 3からA52までに入力してください。番号を入力したら、B列に得点のデータを入力してください。

② 適当な階級を設定します。

データがどの範囲にあるかを調べるために、データの最大値と最小値を求めます。A54とA55に**最大値**、**最小値**と入力しておき、B54とB55にそれぞれの値を計算してください。最大値と最小値を含むように、階級を30から10点きざみで、30～40、40～50、…、90～100の7階級とします。階級の数をいくつにするのが適切かというのは重要な問題ですが、経験的に階級の数はデータ数が50～100のとき6～10、データ数が100～250のとき7～12、データ数が250以上のとき10～20にするとよいことが知られています。

Excelでは、階級の上限值を使って度数を計算します。A59に**階級の上限值**と入力し、A60から下側に順に各階級の上限值を40、50、…、100と入力します。階級の上限值はその階級に含まれ、下限より大きく上限以下の度数が求められます。

③ 度数の計算

メニュー[ツール(T)]→[分析ツール(D)]→[ヒストグラム]を選択します。「ヒストグラム」ボックスが表示されますので、[入力範囲(I)]にデータ範囲のB 3からB52までをドラッグして指定します。同様の方法で、[データ区間(B)]に階級の上限值の範囲A60からA66までを指定します。

次に、出力オプションの[出力先(O)]をクリックし、出力先のセルを指定します。ここでは、出力先のセルをC59にしてください。[OK]をクリックすると、データ区間と頻度が表示されます。データ区間は階級の上限值に、頻度は度数に対応しています。

④ 度数分布表の作成

階級値や相対度数などを計算し、度数分布表を完成させましょう。F58に**試験得点の度数分布表**、F59から順に右側に**下限値**、**上限値**、**階級値**、**度数**、**相対度数**と入力し、F60から下側に順に各階級の下限值30、…、90を、右の列に各階級の上限值40、…、100を入力してください。次に、階

5. 2 代表値と散布度

1次元のデータが得られたとき、度数分布表を作成し、ヒストグラムを描くことはデータの特性を視覚的に把握できるというメリットがあります。しかしながら、データのもつ特徴を数理的に分析したり、他のデータと比較したりする場合には表や図だけでは十分とはいえません。そこで、データの特徴を客観的・数量的に表現する必要があります。

5. 2. 1 代表値

データ全体の中心の位置あるいは代表する位置を表す値を**代表値**といいます。

① **平均値** 平均値(mean)は代表値の中で最もよく使われています。 n 個のデータを x_1, x_2, \dots, x_n とすると、その平均値 \bar{x} は

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

で定義されます。Excel では AVERAGE 関数で計算することができます。

② **中央値** n 個のデータ x_1, x_2, \dots, x_n を小さい方から並べたものを $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ とすると、中央に位置する値のことを**中央値**または**メディアン**といい、 Me で表します。中央値は n が奇数か偶数かによって

$$Me = \begin{cases} x_{\left[\frac{n+1}{2}\right]} & (n \text{ が奇数のとき}) \\ \frac{1}{2} \{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\} & (n \text{ が偶数のとき}) \end{cases}$$

で定義されます。データの中に極端に大きい値や小さい値があるとき、平均値はそれらに大きく影響されますが、中央値は極端なデータには左右されないという性質があります。Excel では関数 MEDIAN で中央値を求めることができます。

③ **最頻値** データの中で最も出現回数が多い値のことを**最頻値**あるいは**モード**といい、 Mo で表します。また、度数分布表では度数が最も多い階級の階級値が最頻値となります。Excel には最頻値を求める関数 MODE があります。

5. 2. 2 散布度

データのばらつきを表す尺度を**散布度**といいます。

- ① **範囲** n 個のデータを小さい方から $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ と並べたとき、最大値と最小値の差

$$R = x_{(n)} - x_{(1)}$$

を**範囲**(range)といいます。範囲はデータ数の増加に伴って大きくなる傾向にあり、特にデータの中に極端に大きい値や小さい値があるとき、それらに大きく影響されるという欠点があります。

- ② **四分位偏差** n 個のデータを小さい方から $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ と並べたとき、 n 個のデータを 4 等分する位置にある値を小さい方から順に Q_1 (第 1 四分位数)、 Q_2 (第 2 四分位数)、 Q_3 (第 3 四分位数) とします。このとき

$$Q = \frac{Q_3 - Q_1}{2}$$

を**四分位偏差**(quartile deviation)といいます。四分位偏差は極端な値をもつデータの影響を受けにくいという利点があります。Excel では関数 QUARTILE で、数値で指定された四分位数を求めることができます。

- ③ **平均偏差** n 個のデータ x_1, x_2, \dots, x_n およびその平均値 \bar{x} に対し

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

を**平均偏差**(mean deviation)といいます。個々のデータ x_i と平均値 \bar{x} との差 $x_i - \bar{x}$ は**偏差**(deviation)とよばれます。

- ④ **分散と標準偏差** 散布度の中で最もよく用いられているのが、分散と標準偏差です。 n 個のデータ x_1, x_2, \dots, x_n およびその平均値 \bar{x} に対し

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

を**分散**(variance)といい、分散の正の平方根

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

を**標準偏差**(standard deviation)といいます。Excel では関数 VAR で分散 s^2 、STDEV で標準偏差 s を計算することができます。

※ 分散と標準偏差の定義として、 s^2 と s の代わりに

$$s'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s' = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

を使うこともあります。Excel では関数 VARP で分散 s'^2 ，STDEV.P で標準偏差 s' を計算することができます。

実習 7 統計関数を使って、成績データの平均値、中央値、範囲、四分位偏差、分散、標準偏差を求めてください。

6 2次元データの解析

これまではテストの点数のように1つの変数についてのデータを扱ってきましたが、ここでは、身長と体重、年齢と血圧、喫煙本数と肺がん発生率、数学の成績と英語の成績など、2つの変数についての観測値が同時に得られるようなデータ(2次元データ)の整理・分析を行い、2つの変数間の関係について考えてみましょう。

6.1 散布図

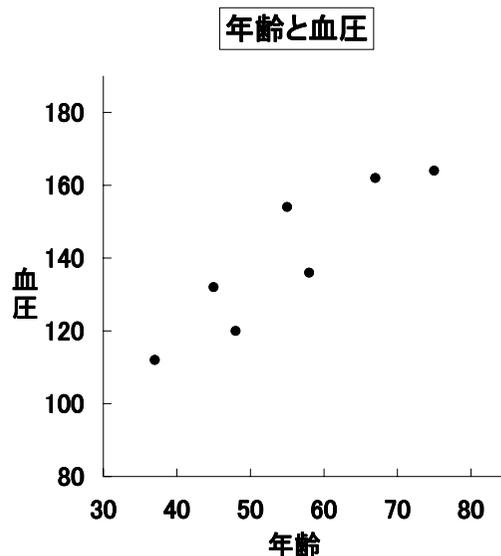
身長と体重のように、2つの変数の観測値が同時に得られたときの n 組の2次元データを $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ で表します。これらの各データを、 x_i を x 座標、 y_i を y 座標として x y 平面上にプロットしたものを**散布図**あるいは**相関図**といいます。散布図を描くと2つの変数間の関係がどのようになっているかを視覚的にとらえることができます。散布図において、2つの変数 x 、 y の一方が増加するとき他方も増加する傾向があるとき、 x と y の間に**正の相関**があるといい、逆に一方が増加するとき他方が減少する傾向があるとき、**負の相関**があるといいます。また、どちらの傾向もみられないとき、**無相関**あるいは**相関がない**といいます。たとえば、身長と体重の間には正の相関があると考えられ、山の標高と気温の間には負の相関があると考えられます。また、散布図において各点の並び方が直線的であるほど、強い相関があるといい、無相関に近いほど、弱い相関があるといいます。

次のデータは7人の女性の年齢と血圧を調べた結果です(杉山(1984))。

被験者	年齢 x	血圧 y
1	37	112
2	45	132
3	48	120
4	55	154
5	58	136
6	67	162
7	75	164

実習8 このデータをセルA1からC8に入力し、以下の手順にしたがって、年齢と血圧の散布図を描いてください。

データのあるB2からC8までの範囲をドラッグし、指定しておきます。標準ツールバーの[グラフウィザード]ボタンをクリックすると、4ページの「グラフウィザード」ボックスが表示されます。1ページ目のグラフの種類で[散布図]→[散布図-値の組を比較します]を選択します。3ページ目のグラフオプションの「タイトルとラベル」で、[グラフタイトル(T)]に**年齢と血圧**、[X/数値軸(A)]に**年齢**、[Y/数値軸(V)]に**血圧**と入力し、[凡例]をクリックして[凡例を表示する(S)]の前のボックスがチェックされていない状態にします。4ページ目のグラフの作成場所で[完了(F)]をクリックすると、散布図が作成されます。



6. 2 相関係数

2つの変数 x, y の間の直線的な関係の強さを表す数値として、**相関係数**(correlation coefficient)がよく用いられます。 n 組の2次元データ (x_i, y_i) ($i=1, 2, \dots, n$) に対し、変数 x, y の平均値、標準偏差をそれぞれ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

とし、 x と y の共分散 (covariance) を

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

とおくと、相関係数 r は

$$r = \frac{s_{xy}}{s_x s_y}$$

で定義されます。相関係数 r は次のような性質があります。

- ① 変数 x と y の間に正の相関があるときは $r > 0$ 、負の相関があるときは $r < 0$ となる。
- ② r は $-1 \leq r \leq 1$ を満たす。
- ③ $|r|$ が 1 に近いほどデータは 1 本の直線の近くに集中する。

実習 9 以下を参考にして、年齢と血圧のデータから相関係数を求めてください。

Excel では統計関数 CORREL(x のデータ範囲, y のデータ範囲) で相関係数を求めることができます。また、分析ツールを使って求めることもできます。メニュー [ツール(T)] → [分析ツール(D)] → [相関] を選択します。「相関」ボックスが表示されますので、[入力範囲(I)] にデータ範囲の B2 から C8 までをドラッグし指定します。出力先のセルを指定して、最後に [OK] をクリックすると

	列 1	列 2
列 1	1	
列 2	r	1

のように表示されます。これは**相関行列**とよばれ、行列の非対角の要素 r が相関係数に相当しま

す。

3つ以上の変数を同時に扱い、各変数間の相関係数を求めるときに便利です。

6. 3 回帰分析

2つの変数 x , y が直線的な傾向を示している場合、 n 組の2次元データ (x_i, y_i) ($i=1, 2, \dots, n$) に最もよく適合した式 $y = \hat{\beta}_0 + \hat{\beta}_1 x$ を求めることを考えます。このために、各データの直線からの“ズレ”を考慮に入れて次の関係式を想定します。

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i=1, 2, \dots, n) \quad (6.1)$$

(6.1)を線形単回帰モデルといい、 β_0, β_1 を回帰係数、 ε_i を誤差項とよびます。 β_0, β_1 の決め方の一つの方法として、各データから直線へおろした鉛直線の長さ、すなわち“ズレ”の2乗和

$$f(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \{y_i - (\beta_0 + \beta_1 x_i)\}^2$$

を最小にする β_0, β_1 を求める方法が知られ、これを最小2乗法といいます。 $f(\beta_0, \beta_1)$ を最小にする β_0, β_1 の値を各々 $\hat{\beta}_0, \hat{\beta}_1$ とすると、これらは

$$\begin{aligned} \frac{\partial f}{\partial \beta_0} &= -2 \sum_{i=1}^n \{y_i - (\beta_0 + \beta_1 x_i)\} = 0 \\ \frac{\partial f}{\partial \beta_1} &= -2 \sum_{i=1}^n \{y_i - (\beta_0 + \beta_1 x_i)\} x_i = 0 \end{aligned}$$

の解として与えられ、 \bar{x}, \bar{y} を変数 x, y の平均値とすると

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

と表すことができます。この $\hat{\beta}_0, \hat{\beta}_1$ を係数に持つ式

$$y = \hat{\beta}_0 + \hat{\beta}_1 x \quad (6.2)$$

を回帰式あるいは回帰直線とよびます。さらに(6.2)式における x を説明変数あるいは独立変数、これに対して y を目的変数あるいは従属変数といいます。

実習 10 以下を参考にして、年齢と血圧のデータから回帰式の係数 $\hat{\beta}_0, \hat{\beta}_1$ を求めてください。

Excel には回帰式 $y = \hat{\beta}_0 + \hat{\beta}_1 x$ の傾きと切片を求める統計関数 SLOPE(y のデータ範囲, x の

ータ範囲)と INTERCEPT(y のデータ範囲, x のデータ範囲)が用意されています。また, 分析ツールを使って求めることもできます。メニュー[ツール(T)]→[分析ツール(T)]→[回帰分析]を選択します。「回帰分析」ボックスが表示されますので, [入力 Y 範囲(Y)]に C 2 から C 8 までを, [入力 X 範囲(X)]に B 2 から B 8 までをドラッグし指定します。出力先のセルを指定して[OK]をクリックすると, 回帰分析の結果が出力されます。3つの表のうち一番下の表で, **切片**と表示されているセルの右側が $\hat{\beta}_0$, **X 値 1**と表示されているセルの右側が $\hat{\beta}_1$ に相当します。

実習 11 回帰式を用いて, 50 歳の女性の血圧値を予測してください。また, 年齢と血圧の散布図の中に回帰式を描き入れて, データへのあてはまりの様子を確かめてください。(散布図のデータ上を右クリックすると, ショートカットメニューが表示されます。[近似曲線の追加(R)]→[線形近似(L)]を選択してください。回帰直線上で右クリックし, [近似曲線の書式設定(O)]を選択すると直線の書式が変更できますので, [オプション]→[予測]の[前方補外]と[後方補外]を適当に設定して見やすくしてください。)

7 グラフの作成

Excel ではデータに基づいて各種のグラフを作成することができます。グラフを作成するには, グラフの元になるデータ範囲を選択し, 標準ツールバーの[グラフウィザード]ボタンをクリックします。グラフにはあらかじめ標準グラフ 14 種類とユーザー設定グラフ 20 種類が用意されており, データの内容によって目的に合ったグラフを選択することが可能です。また, Excel では基本的な関数のグラフを容易に作成することができます。例えば, $y = x^2$ (定義域 $[-3, 3]$) のグラフを描いてみましょう。

セル A 1 に x , B 1 に y と入力してください。次に, A 2 から下側に定義域となる -3 から 3 まで 0.1 ずつ増加させた値を入力します。B 2 に $=A2^2$ と入力し, これをコピーしてすべての x の値に対する y の値を求めておきます。 x と y の範囲をドラッグし, 標準ツールバーの[グラフウィザード]ボタン→[散布図]→[データポイントを平滑線でつないだマーカーなしの散布図]を選択すれば, 関数 $y = x^2$ のグラフを作成することができます。

実習 12 次の関数のグラフを描いてください。

- (1) $y = x, y = x^2, y = x^3$ (定義域 $[-1, 1]$, 同時に)
- (2) $y = \sin x, y = \cos x$ (定義域 $[-7, 7]$, 同時に)
- (3) $y = \arcsin x, y = \arccos x$ (定義域 $[-1, 1]$, 同時に)
- (4) $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (定義域 $[-3, 3]$)

8 行列の計算

Excel では行列の和、差、実数倍などの計算を実行することができます。また、行列の転置、積、逆行列、行列式を求めるための関数が用意されています。行列計算を行う場合、まず初めに計算の結果が何行何列になるかを確認し、出力先をドラッグして範囲指定しておきます。結果を出力する際は **Ctrl** キーと **Shift** キーを押しながら **Enter** キーを押します。

和と差 =と入力し、一方の行列をドラッグします。次に、和を求める場合は+を、差を求める場合は-を入力して、もう一方の行列をドラッグします。

実数倍 = 実数倍する数 * と入力し、実数倍したい行列をドラッグします。

転置行列 [関数貼り付け] ボタン f_x → [検索/行列] → [TRANSPOSE] を選択します。[配列] に行列データをドラッグして指定します。

積 [関数貼り付け] ボタン f_x → [数学/三角] → [MMULT] を選択します。[配列 1] と [配列 2] に 2 つの行列データをドラッグして指定します。

逆行列 [関数貼り付け] ボタン f_x → [数学/三角] → [MINVERSE] を選択します。[配列] に行列データをドラッグして指定します。

行列式 [関数貼り付け] ボタン f_x → [数学/三角] → [MDETERM] を選択します。[配列] に行列データをドラッグして指定します。

実習 13 次の計算をしてください。

$$(1) \begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 5 & -2 \\ -3 & 4 \end{pmatrix} \quad (2) 2 \begin{pmatrix} 1 & 5 & -3 \\ -1 & 2 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 2 \\ 2 & -3 & 1 \end{pmatrix}$$

$$(3) \begin{pmatrix} 3 & 2 \\ 6 & -1 \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -3 & 2 \end{pmatrix} \quad (4) \begin{pmatrix} 1 \\ 5 \end{pmatrix} (-3 \ 2) \quad (5) (1 \ 2) \begin{pmatrix} -2 & 3 \\ 5 & -6 \end{pmatrix}$$

実習 14 $A = \begin{pmatrix} 3 & 2 \\ -2 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$ に対し、次の等式が成り立つことを示してください。

ただし、 ${}^t X$, X^{-1} , $|X|$ はそれぞれ行列 X の転置行列、逆行列、行列式を表します。

$$(1) {}^t(AB) = {}^t B {}^t A \quad (2) (AB)^{-1} = B^{-1} A^{-1} \quad (3) ({}^t A)^{-1} = {}^t(A^{-1})$$

$$(4) |AB| = |A||B| \quad (5) |{}^t A| = |A|$$

9 成績処理について

5章の成績データを使用します。項目名(生徒番号, 得点)を含めて, A 2 から B52 までを新しいワークシートの A 1 から B51 へコピーしてください。

9. 1 順位づけ

データに順位をつけるための関数として, RANK 関数が用意されています。RANK 関数は[関数貼り付け]ボタン  → [統計] → [RANK] を選択します。一般に =RANK(数値, 範囲, 順序)のように用いられ, 数値で指定した値が, 指定した範囲の中で何番目にあたるかを求めることができます。順序には 0 か 1 を指定します。「0」は降順に並べたときの順位, 「1」は昇順に並べたときの順位を求めるときに指定します。ただし「0」は省略が可能です。C 1 に **順位**, C 2 に =RANK(B2, \$B\$2:\$B\$51, 0)と入力し, これを C 5 1 までコピーすればすべての生徒番号に対して得点の大きい順に順位をつけることができます。

9. 2 データの並べ替え

データ(数値)を小さい順あるいは大きい順に並べ替えるには, 並べ替えを実行したい項目名をクリックして, データを小さい順に並べ替える場合には標準ツールバーの[昇順で並べ替え]ボタンを, 大きい順に並べ替える場合には[降順で並べ替え]ボタンをクリックします。

実習 15 得点の高い順に並べ替えを行ってください。次に, 生徒番号を昇順に並べ替えてください。

9. 3 データの抽出

データの抽出は基本的に**オートフィルタ**を使って行うことができます。リスト内(項目名とデータ範囲)の任意のセルをクリックし, メニューの[データ(D)]→[フィルタ(F)]→[オートフィルタ(F)]を選択します。項目名の右に下向きの矢印が現れますので, その矢印をクリックし, [(オプション)]を選択します。「オートフィルタオプション」ボックスが開きますので, [抽出条件の指定]を行います。左側のボックスに数値を入力し, 右側の[と等しい]が表示されているボックスの右側の矢印をクリックして抽出条件を選択します。最後に[OK]をクリックすれば, 指定した条件を満足するデータが抽出されます。抽出を解除するには, 項目名の右にある矢印をクリックし, [(すべて)]を選択します。また, オートフィルタを解除するには, [データ(D)]→[フィルタ(F)]→[オートフィルタ(F)]をクリックします。

実習 16 得点が 60 点以上のデータを抽出してください。抽出が終わったら、オートフィルタを解除してください。

9. 4 成績の評価

例えば、80 点以上は A、70 点以上 80 点未満は B、60 点以上 70 点未満は C、60 点未満は D の評価を、データの入力されているシート D 列に表示してみましょう。シートの空いているところ、例えばセル F 1 に**評価基準**、F 2 と F 3 に**得点**、**評価**と入力し、G 2 から右側に 0、60、70、80、G 3 から右側に D、C、B、A と入力します。D 2 に =HLOOKUP(B2, \$G\$2:\$J\$3, 2, TRUE) と入力し、これを D51 までコピーすればすべての評価を求めることができます。

参考 HLOOKUP 関数は [関数貼り付け] ボタン  → [検索/行列] → [HLOOKUP] を選択します。一般に =HLOOKUP(検索値, 範囲, 行番号, 検索の型) のように用いられ、指定したテーブルまたは配列の上端行で特定の値を検索し、その値と同じ値が入力されている行と列の値を返します。HLOOKUP 関数は、比較する値がデータテーブルの上端行に入力され、その位置から指定された行だけ下にある値を取り出す場合に使用します。なお、検索の型には検索値と完全に一致する値だけを検索するか(FALSE)、その近似値を含めて検索するか(TRUE)を指定します。

実習 17 75 点以上は A、60 点以上 75 点未満は B、50 点以上 60 点未満は C、50 点未満は D の評価を E 列に表示してください。

10 おわりに

本報告では、表計算ソフト Excel の基本操作から多彩な機能や特徴の一部を紹介し、Excel が統計データの処理・解析や数学・情報教育などへ幅広く活用できることを述べました。数百種類の関数が用意されているだけでなく、作業を効率的に実行できるマクロ機能やユーザーのニーズにあった関数を作成できるユーザー定義関数などにより、さまざまな計算・分析を実行することが可能です。また通常の関数では処理しきれない、より高度で複雑な計算・分析を行う場合には Excel VBA(Visual Basic for Application)という言語を利用して独自のプログラムを作成することもできます。近年 VBA に関する書籍が数多く出版されており、情報教育等への利用も大いに期待されます。

参考文献

- [1] 金 明哲, 中村永友, 山田智哉(2003):データ解析の基礎, ムイスリ出版.
- [2] 杉山 高一(1984):統計学入門, 絢文社.
- [3] 縄田 和満(2000):Excelによる統計入門, 朝倉書店.
- [4] 馬場 裕(2002):確率と統計の基礎・基本, 牧野書店.

(Received Mar. 31, 2005)