

第9章 データ処理と表計算ソフトウェア

本章では、一般に「表計算ソフトウェア」と呼ばれているアプリケーションソフトウェアの機能の紹介と使い方について解説をする。このガイドブックを見ながら実際にPCを操作し、習熟しておいてほしい。本校では、質問紙調査（アンケート）の集計や実験データの分析、グラフ作成などにMicrosoft社のExcelを使用しているので、このソフトウェアについての解説となる。

なお、このソフトウェアには様々な機能があるが、すべての機能について習熟しておく必要はない。操作方法を忘れたり、わからなかったりするときには友達に聞か、インターネットの検索機能を使って操作方法を解説しているサイトを閲覧しながら操作をしていくとよい。例えば、標準偏差を算出する関数がわからなければ、「Excel 関数 標準偏差」などといくつかのキーワードをスペース（空白）で区切って検索ボックスに入力すると、Excelで使える関数について解説した多くのサイトが表示される。中には、動画で懇切丁寧に解説しているサイトもある。周囲に友達がいる場合は、「聞きまくる」のが習熟の早道であろう。逆に、ソフトウェアの操作が得意な人は、友達に尋ねられたときには惜しみなく教えてあげてほしい。

本章では、Excelの基本的な機能である関数の使い方や、グラフの作成などについて説明する。

9.1 セル、シート、ブック

早速だが、PCのデスクトップにある緑色の[X]のアイコンをダブルクリック（マウスの左ボタンを「カチカチッ」とすばやく2回押すこと）してExcelを立ち上げる。続いて「空白のブック」をクリック（マウスの左ボタンを1回だけ「カチッ」と押すこと）する。縦横の罫線が入った真っ白な入力画面が表示される。一つ一つのマス目のことを、「セル」といい、今見えている画面を「シート」という。実験データなどは、この「セル」の中に一つ一つ入力していくことになる。余談になるが、このシートを見ると、あたかも中学校の理科の教科書に掲載されていた「タマネギの細胞」を彷彿とさせる。生物では「Cell：セル」といえば細胞のことを言い、物理では「電池」のことを言うことが多い。シートの方角を見ると、「Sheet1」と表示されたタブが見える。「シート」は「Sheet1」のすぐ右にある⊕（丸に十字）をクリックすることにより、自由に増やすことができる。シートの名前の変更も、タブの上で右クリックし[名前の変更]を選択することにより自由にできる。

文字や数値の入力については、本書の題6章（論文とポスターの作成）にある表2を参照しながら、入力していこう。全角・半角（A・A, a・a, ア・ア, 3・3）の違いなどに細心の注意を払いながら入力してみよう。なお、計算するために入力する数値は、半角（例：123）で入力する。全角の数字（例：1 2 3）で入力すると、「文字列」と認識されるので注意を要する。

入力モード（「日本語入力モード」か「英数字入力モード」か「半角入力モード」など）は、次の図1のように、画面右下隅に「あ」と表示されているか「A」と表

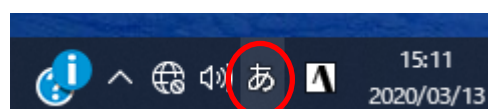


図1 日本語入力モードの確認

示されているか「半」と表示されているかで確認できる。「あ」と表示されていたら、全角入力が可能で、このモードで入力をした後で半角文字に変更することも可能（スペースキーを押して半角に変換するなど）である。

残念だが、カタカナの半角文字「ｱｲｳｴ」はあるが、ひらがなについては全角文字「あいいうえお」だけで半角文字はない。

セルには、「番地」があり、図2の黄色のセルは、セル番地「A2」で、青色の四つの複数セルは、「B4:C5」と表記する。セルを選択（Select）するときには、一つのセルであれば、セルの真ん中付近をクリックすればよい。図2の青い領域のように、複数セルを選択する場合は、まず先頭のB4セルの真ん中付近で左ボタンを押さえ、押さえたまま斜め下方向にドラッグしてC5セルの真ん中付近でボタンを離す。選択されたセルは、灰色の網掛けの状態になる。

	A	B	C
1			
2			
3			
4			
5			
6			

図2 セル番地

9.2 平均と合計、関数

平均と合計の算出の方法を紹介する。ここでは、「関数」と呼ばれる便利な機能の使い方に習熟してほしい。まずは、1から10までの連続した数値をB2:B11の範囲のB列に入力する。Enterキーを押すごとにセルが一つずつ下がっていく。10個のデータであれば手入力でもさほど手間はかからないが、100までとなると大変である。このようなとき、連続したデータを機械的に入力できる便利な機能「オートフィル」（「自動的に満たす」の意味）があるので紹介する。まず、

① 二つのセルの選択

B21とB3にそれぞれ1と2を入力し、B2:B3の範囲を選択する。

	A	B
1		
2		1
3		2
4		
5		

この状態で、緑の枠線の右下隅の小さな四角■のちょうど真上にカーソルを合わせるとカーソルの形が黒十字+に変わる。

② ドラッグ

カーソルが常に黒十字+の状態を保ち、左ボタンを押さえ、そのまま下方にドラッグする。「10」と表示されたところでボタンを離す。

成功すれば、B2:B11の範囲に、1から10までの連続した数値が入力できている。

慣れるまで、何度か試みよう。

【やってみよう！ オートフィルの便利な機能】

- ・二つのセルに、1, 3と入力してオートフィルを実行すると、どのような数値の並びになるかやってみよう。
- ・二つのセルに、10, 20と入力してオートフィルを実行すると、どのような数値の並びになるかやってみよう。
- ・二つのセルに、「月曜日」「火曜日」と入力してオートフィルを実行すると、どのような文字が表示されるかやってみよう。

B12 セルに平均値を表示させる。

平均値を算出するための関数 (Function) を B12 に埋め込むと、平均値が表示される。まず、B12 に半角= (イコール) の記号を入力する。半角イコール「=」を入力することで、Excel は関数が入力されること認識する。続いて平均を意味する Average をタイプする。関数の入力は大文字・小文字 (A・a) の区別はないので、どちらで入力し

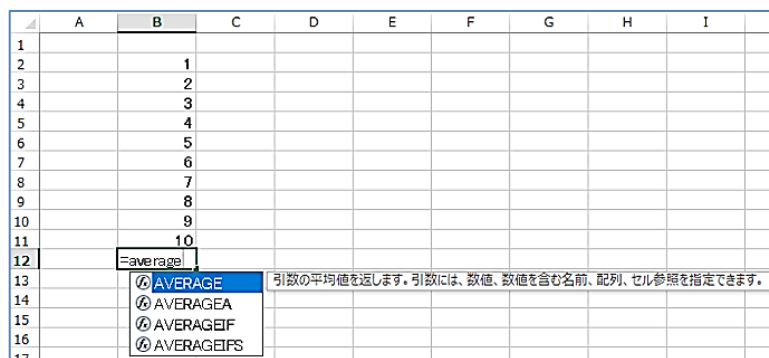


図3 関数の入力

てもかまわない。ただし、全角文字は使えないので必ず半角で入力すること。入力の途中で、図3のようにプルダウンメニューが表れ、関数の候補が表示されるので、今回の場合は、「AVERAGE」をダブルクリックすると自動的に関数が入力できる。「=AVERAGE(」が入力された状態を確認した後、範囲を指定する。関数が表示されている状態で、B2 から B11 をドラッグすると、自動的に範囲が入力され、「=AVERAGE(B2:B11)」のように表示される。ここで Enter キーを押すと B12 に「5.5」が表示される。なお、セル範囲の B2:B11 は、手動で入力してもかまわない。

続いて B13 に合計を表示させてみよう。使用する関数は、SUM 関数 (Summation: 合計の略) である。B13 に「=SUM(B2:B11)」となった状態で Enter キーを押すと、合計の「55」が表示される。

このように関数の機能は大変便利である。日常でもよく使う平均や合計の関数以外にも、三角関数など多くの関数が用意されている。

9.3 代表値と度数分布表・ヒストグラム

個々では代表値として「平均値 (Average)」「中央値 (Median)」「最頻値 (Mode)」の三つを紹介する。これらの英語の綴りがそのまま Excel の関数となるので便利である。

ここで、平均値と中央値にずいぶんひらきがある例として、金融広報中央委員会が毎年行なっている「家計の金融行動に関する世論調査」の結果を紹介しよう。ここで言う「金融資産」とは、預貯金などの金融商品を指し、現金、土地、住宅、貴金属は含まない。ある年の調査結果によると、各世帯の金融資産の平均値は「1,078 万円」で、資産が大きい世帯があるため、平均値を押し上げている。金融資産が、まったくない世帯も 30%以上あるとのことである。そのため、金額の低い世帯から高い世帯へと順に並べ、ちょうど中央に当たる中央値でみると、1 世帯の金融資産は「400 万円」になる。この中央値は、平均値の半分弱であり、ずいぶんと乖離している。

このように、同じ代表値でも、データの分布によってはずいぶんとひらきが出てくる。

次に、具体的な事例を基に、Excel で分析してみよう。

(例①) あるクラスの「一か月の読書時間」について、電車・バス通学者のA班と徒歩・自転車通学のB班に分けて調べる。どちらのグループの読書時間がより優れているといえるだろうか。

A 班 20 人(単位:時)										
3	10	7	14	5	9	15	0	13	18	
0	8	11	10	15	19	6	23	9	5	

(B3:B22 にA班のデータを入力)

B 班 19人(単位:時)										
26	0	15	1	5	4	2	10	9	16	
20	23	4	6	8	30	7	3	1		

(C3:C21 にB班のデータを入力)

①平均値

平均値は、データの値の総和を総数で割ったものである。

数式で表すと $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ となる。この式は、平均値 $\bar{x} = (x_1 + x_2 + \dots + x_n) / n$ を意味している。 x の値の平均値 \bar{x} は、「エックス・バー」と読む。ギリシャ文字の Σ (大文字のシグマ) は和を表し、1 番目の値から n 番目の値までを足し合わせることを意味している。

(例 1) A 班の平均値を求める。

【関数式の定義】

=AVERAGE (範囲 1, 範囲 2, ...)

※ 範囲はセル番地で表記し、複数範囲を使用する場合はカンマ(,)で区切る。

【A 班の平均値を求める関数式】

=AVERAGE (B3:B22)

同様にして B 班の平均値を求めてみよう。

②中央値(メジアン)

すべてのデータの値を小さい順に並べたとき、中央の順位にくる値のことである。

※ データの個数が $2n$ 個 (偶数) のとき、 n 番目と $n+1$ 番目のデータの平均値を中央値とする

(例 2) A 班の中央値を手計算と Excel で求める。

=MEDIAN (範囲 1, 範囲 2, ...)

※ 範囲はセル番地で表記し、複数範囲を使用する場合はカンマ(,)で区切る。

【A 班の中央値を求める関数式】

=MEDIAN(B3:B22)

同様にして B 班の中央値を Excel で求めてみよう。

③最頻値(モード)

データを度数分布表に整理したとき、度数が最も大きい階級の階級値のことである。

【Excel を使った度数分布表の作り方】

次の手順で Frequency 関数を使って度数分布表を作成してみよう。

- (1) E 列に区間を作成する。
 ※ 3 から 4 刻みにする (E3~E9 に昇順で入力)
 続いて次の手順で、F 列に度数の関数を入力する。
 * (2) F3~F9 を範囲選択し、FREQUENCY 関数を呼び出す。
 (複数セルが選択された状態のまま、=FREQUENCY を入力)
 * (3) データ配列を B3~B22、区間配列を E3~E9 とし、
Ctrl と Shift を押しながら Enter キーを押す。

A 班の度数分布表	
区間	度数
3	3
7	4
11	6
15	4
19	2
23	1
30	0

Frequency は頻度を意味する。物理では振動数を意味することが多い。

度数分布表の「区間」の「3」は「3 以下」を意味し、区間の列に入力する数値は、小さい方から大きい方へと昇順にしておく必要がある。

上記の操作はやや高度なので忘れた場合は、インターネットで検索して手順を調べればよい。

(例 3) A 班の最頻値を求める。

度数分布表より、度数が最も大きい階級は 7 時間より大きく 11 時間以下のときである。(7 時間を超え、11 時間以下に 6 人)

よって、最頻値はこの階級の階級値を求めて、9 時間となる。

※階級値・・・階級における真ん中の値 7~11 の場合、階級値は 9

B 班の度数分布表を Excel で作成し、最頻値を求めてみよう。

(番外編)MODE 関数で求めることもできる。A 班、B 班の最頻値を、MODE 関数を使って求めよ。

【関数式の定義】

=MODE (範囲 1, 範囲 2, ...) (MODE. SNGL でも同様の結果が得られる。)

※ 範囲はセル番地で表記し、複数範囲を使用する場合はカンマ(,)で区切る。

【A 班の最頻値を求める関数式】

=MODE. SNGL (B3:B22)

代表値それぞれの特徴は次のとおりである。

平均値	全てのデータを平等に扱うが、外れ値に弱い
中央値	外れ値に強いが、全てのデータを平等に扱うわけではない
最頻値	外れ値に強いが、極端な分布に対応できないことがある

データ分析は、平均値だけではなく、さまざまな統計値、分布のグラフを用いて比較することが重要である。

9.4 グラフの作成

前節の「A班の度数分布表」のような形でデータを入力しておくで、比較的簡単に様々なグラフを作成することができる。この表全体（項目名等を含む）を選択した後、[挿入]-[グラフ]から、最も適したグラフを選択することができる。前節の度数分布表から「ヒストグラム」を作成してみよう。

なお、作成した直後は、ラベルと単位などが表示されていない場合が多い。論文等に掲載する正式なグラフにする場合は、グラフを選択し、すぐ右上に表示される \boxplus （グラフ要素）をクリックしてラベルやなど、必要なグラフ要素を追加していくことになる。自然科学系では、散布図をよく使うことになる。この場合、「近似曲線の追加」で近似曲線や、その数式を表示させることもできる。

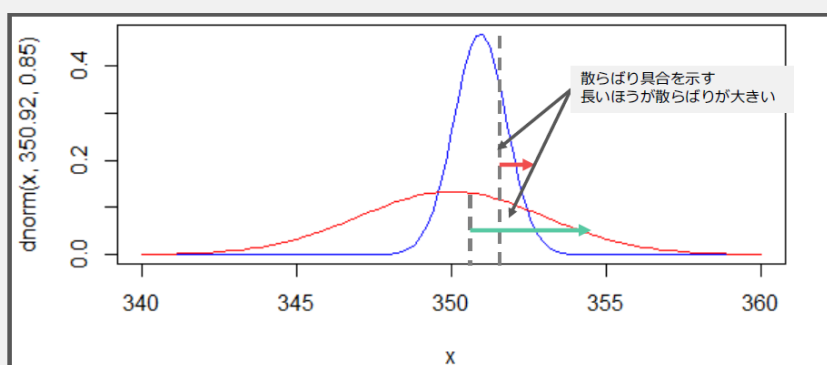
棒グラフや円グラフなど、様々なグラフが用意されているので、いろいろと試してみよう。なお、グラフを作成するとカラーになるが、例えば棒グラフの場合は、棒の部分をクリックすることにより、色を変えたりモノクロ（白黒）にしたり、様々なパターン（網掛け、斜線など）に変えたりすることも可能である。

9.5 分散と標準偏差

平均値が同じでも、データの散らばり具合がずいぶんと異なる場合がある。この「データの散らばり具合」の目安となる数値を「分散」(s^2)、「標準偏差」(s)と言う。標準偏差は分散の平方根である。逆に言うと標準偏差の2乗が分散となる。散らばり具合を視覚化した次の図にある二つのグラフを比較すると、平均値付近に集中している青色のグラフの場合は標準偏差が小さく、なだらかな赤色のグラフの場合は標準偏差が大きいと言える。正規分布を仮定すると、 $\pm s$ の間にデータの約68%が入り、 $\pm 2s$ の間には95%が入っていることになる。

分散と標準偏差

散らばり具合を視覚化すると



分散の定義式は次のようになる。

$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ と表され、具体的な計算は次の式で表される。

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

標準偏差は分散の平方根なので

$$s = \sqrt{s^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

幸いなことに、Excel を使うと、上の式の通りに四則演算をしなくても、分散 (Variance) の関数 (=VAR) と標準偏差 (Standard Deviation) の関数 (=STDEV) を使えば、すぐに答えが得られる。

練習のために、次の例 4 を Excel でやってみよう。

(例 4) ある携帯音楽プレイヤーを充電してからの連続使用時間を 10 回調べた結果を以下に示す。

連続使用時間(単位:時間)

24	21	26	23	26	20	22	26	22	30
----	----	----	----	----	----	----	----	----	----

【前半 5 回分のデータを使って分散を求める】

(1) データの平均値を求める。

=AVERAGE (B3:B7)

(2) 各データの偏差を求めて 2 乗する。

※ 偏差・・・それぞれのデータと平均値との差

(C3)=\$B\$8-C3 (D3)=c3^2

[\$は絶対セル番地で、コピーや移動しても不変、^2 は 2 乗を表す]

(3) 分散 s^2 を求める

=SUM (D3:D7) /COUNT (B3:B7) [COUNT 関数でデータ数を求める]

※ 分散は、計算の過程で数値を 2 乗するので単位が変わることに注意。

練習のために、10 回分のデータを使って分散を求めてみよう。

(簡単編)VAR 関数で求めることができる。5, 10 回分の分散を、VAR 関数を使って求めてみよう。

【関数式の定義】

=VAR. P (範囲 1, 範囲 2, ...) (VAR. P は母分散, VAR. S は標本分散)

【5 回分の分散を求める関数式】

=VAR. P (B3:B7)

なお、P は Population (母集団)、S は Sample (標本) を意味する。

(簡単編)STDEV 関数で標準偏差求めることができる。5, 10 回分の標準偏差を、STDEV 関数を使って求めてみよう。

【関数式の定義】

=STDEV. P (範囲 1, 範囲 2, . . .) (STDEV. P は母集団標準偏差, STDEV. S は標本標準偏差)

【5 回分の標準偏差を求める関数式】

=STDEV. P (B3:B7)

実験データが得られたら、平均値だけでなく、標準偏差も併記しておくでデータの散らばり具合がよく分かってよい。

9.6 相関係数

二つのデータ (x, y) のうち、一方が増加すると他方も増加する傾向にある場合、正の相関があるという。また、一方が増加すると他方が減少する傾向にある場合、負の相関があるという。どちらの傾向もない場合は、相関がないという。

【相関係数】

二つのデータ x, y の共分散を s_{xy} とし、 x と y の標準偏差を s_x, s_y とすると、相関係数 r は次の式で表される。

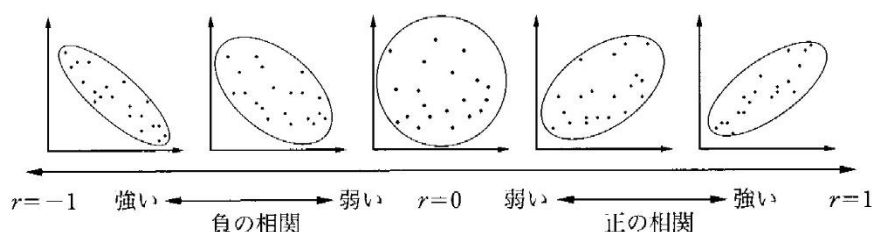
$$r = \frac{s_{xy}}{s_x s_y}$$

ここで、共分散 s_{xy} は次の式で表される。 $s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

相関係数 r の値は次の不等式が成り立つ。 $-1 \leq r \leq 1$

特に、正の相関が強いほど r は 1 に近づき、負の相関が強いほど r は -1 に近づく。

$r=0$ のとき	相関なし
$0 < r < 0.4$ 、 $-0.4 < r < 0$ のとき	非常に弱い相関
$0 < r < 0.6$ 、 $-0.6 < r < 0$ のとき	やや相関がある
$r > 0.6$ 、 $r < -0.6$ のとき	強い相関がある



9.7 検定

新聞などで、「学力調査の結果、有意水準1%で有意差があった」「有意水準5%で有意差があった」などの表現を目にすることがある。例えば、集団の平均値に3点の差があったとしても、この差が本当に意味のある差なのか、たまたま偶然に差が出ただけで、次回テストをすると逆転することもあるのかについては、「3点」という数値だけでは判断できない。集団の大きさや分布の状況によって同じ「3点」でも意味合いが異なってくる。そこで「検定」を行う必要が生じる。ここで例に挙げた「1%水準」「5%水準」は偶然に3点の偏りが生じる確率を意味すると考えると、「1%」だと、偶然に差が出た確率が1%しかないということになる。よって、数値の小さい「1%」の方が、より高い確率で差が付いていると言える。

一般的によく使われる検定としてt検定がある。異なる集団の平均値の差に本当に意味があるのか、また、事前事後のテスト結果で本当に伸びが見られるのか、などについて検定にかけた結果を提示すると、ただ平均点だけを提示するよりも説得力が出てくる。

統計的な問題解決の特徴として、 x と y （縦軸と横軸）の二つの変数だけでなく、三つ以上の変数を扱い、これらの変数の結びつき（関連）の強弱を調べるということが挙げられる。「多変量解析」と呼ばれる解析方法である。どのような解析方法が適しているか、事前にしっかり統計の教員とよく相談しておくとうい。

具体的な検定の方法については、必要に応じて統計の教員に相談したり、インターネットや図書館で調べたりしながら習得してほしい。また、先輩たちの事例も参考にするとよい。