

付録 D

電卓とコンピュータを活用する

D.1 電卓で統計計算

家庭用の 300 円ぐらいの電卓でも，簡単な統計の計算に活用することができる．ただし，メモリーキー $\boxed{M+}$ \boxed{MR} が備わっているものでないといけない．ルートキー $\boxed{\sqrt{}}$ はあった方が便利だが，なくても平方根の計算は難しくない（後述）．

基本的な使い方

■メモリーキーの意味

$\boxed{M+}$ メモリーに加算 (memory plus)

\boxed{MR} メモリーを呼び出す (memory recall)

\boxed{MC} メモリーをクリア (memory clear)

\boxed{MRC} メモリー呼び出し，二度押しでクリア (機種によって \boxed{MR} , \boxed{MC} の代わりにある)

■メモリーを使って次々に加算する

加算を繰り返していくときには， $\boxed{+}$ キーよりもメモリー機能を活用したほうが効率がよい．まずは単純な足し算でメモリの使い方を覚えよう．

計算式: $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$

キー操作: $1 \boxed{M+} 2 \boxed{M+} 3 \boxed{M+} \dots 10 \boxed{M+} \boxed{MR} \Rightarrow 55$

結果を消去するには， \boxed{MC} でクリアする．

メモリー機能は次の積和の計算でさらに威力を発揮する．

計算式: $23.5 \times 3 + 41.2 \times 5 + 51.0 \times 6 + 13.5 \times 2$

キー操作: $23.5 \boxed{\times} 3 \boxed{=} \boxed{M+} 41.2 \boxed{\times} 5 \boxed{=} \boxed{M+} 51.0 \boxed{\times} 6 \boxed{=} \boxed{M+} 13.5 \boxed{\times} 2 \boxed{=} \boxed{M+} \boxed{MR} \Rightarrow 609.5$

■2乗を計算

計算式: 12^2 キー操作: 12 $\Rightarrow 144$

■2乗の和を求める

計算式: $1^2 + 2^2 + 3^2 + \dots + 10^2$ キー操作: 1 2 3 ... 10 $\Rightarrow 385$

■平方根

2乗の繰り返し計算で平方根を追い詰めて*1, $\sqrt{30}$ を有効数字3桁まで求めてみる。ドン臭いやり方に見えるが、回数に比例して有効数字が増え、効率はよい。

- $5^2 = 25$, $6^2 = 36$ だから, 5.5 を試す. $\Rightarrow 30.25$ ほんの少し大きすぎ.
- 5.4 を試す. $\Rightarrow 29.16$ 戻し過ぎ.
- 5.45 でどうだ? $\Rightarrow 29.7025$ ちょっと足りない.
- 5.47 でどうかな? $\Rightarrow 29.9209$ まだ足りない.
- 5.48 でどうよ? $\Rightarrow 30.0304$ 行き過ぎたから戻すか.
- 5.475 ならどうだ? $\Rightarrow 29.975626$ 足りない.
- つまり, 5.475 と 5.480 の間になる.
- 5.475 を四捨五入して, 5.48 でオーケー!

生データからの統計計算

■平均

データ: 23.5, 41.2, 50.1, 32.3

キー操作 23.5 41.2 50.1 32.3 4 $\Rightarrow 36.775$

■分散

分散が2乗の平均 - 平均の2乗 で計算できることを利用する。

キー操作: 2乗の平均: 23.5 41.2 ... 4 $\Rightarrow 1450.7475$ メモしておく

平均の2乗から2乗の平均を引く:

36.775 1450.7475 $\Rightarrow -98.347$

符号反転:

符号が反対になっているので, 98.35 が分散。

*1 方程式の解を繰り返し計算で求めるニュートン・ラフソン法 (ニュートン法) という計算手法の原理だ。

標準偏差：

さらに標準偏差を求めるには、平方根をとって $\sqrt{98.35} = 9.92$

度数分布から平均と分散を求める

データ：	階級値	5	15	25	35	45
	度数	2	4	7	4	3
	累積度数	2	6	13	17	20

■平均

MC

5 2 M+

15 4 M+

...

45 3 M+

MR 20 ⇒ 26

■分散

まず 2 乗の平均を求める

5 2 M+

15 4 M+

...

45 3 M+

MR 20 ⇒ 815

次にこの値から平均の 2 乗を引くのだが、次のように順序を反対にしたほうが電卓で計算しやすい。

26 815

⇒ -139

負号を外して、分散 = 139

標準偏差は $\sqrt{139} = 11.79$

D.2 スプレッドシートで統計計算

統計計算のための機能は、Microsoft の Excel などの表計算ソフトにひと通り備わっている。これらを利用して第 1 章で扱った計算を実際に理解するために、数式に沿って一から計算する方法を示す。

CSV ファイルでデータ入力

第 1 章に出てきた表 1.1(p.5) のデータをスプレッドシートに入力するものとする。このとき、セルに直接打ち込むよりも、エディタで別のファイルに予め打ち込んでから作業するほうが便利な事が多いので、そのやり方を説明する。

データをファイルに入力して保存した後、表計算ソフトに読み込ませる。手順は次の通りである。

- 適当なエディタを開く。Windows なら メモ帳、Mac なら テキストエディットが使える。^{*2}
- 数値を半角で 1 行に 1 つずつ入力して改行する。
- 最後のデータの後も改行を忘れずに入れる
- 保存する。ファイル名はなんでもよいが、拡張子は `.txt` としておく
- 保存したファイルの拡張子を `.csv` に変更する（警告が出るが気にせず変更する）
- ファイルのアイコンをダブルクリックする
- 表計算ソフトが立ち上がって、右図のように A 列の A1 から A100 までのセルにデータが入っている。

	A	B	C
1	43.6		
2	45.2		
3	45.4		
4	45.8		
5	47.2		
...			
99	64.2		
100	64.6		

このやり方をとるのは、ふつうのデータ入力では、ウェブ画面や電子ファイルからデータを取り込むことが多いからである。その場合、コピー&ペーストでテキストファイルにデータを書き込むことで、正確な入力と省力化が可能になる。印刷されたデータを使うのであれば、上の手間を掛けずにデータをセルに直接入力するだけでも構わないのだが、練習のためにこの方式でやってみよう。

^{*2} Mac のテキストエディットの初期設定ではリッチテキスト形式になっているので、「環境設定」メニューを開き、「標準テキスト」をチェックして、標準テキストに変更しておくこと。

平均を求める

A101 のセルに `=SUM(A1:A100)/100` と入力する．次の計算を実行したことになり，平均値が表示される．

$$\mu = (43.6 + 45.2 + \cdots + 64.5)/100$$

あるいは `=AVERAGE(A1:A100)` としてもよい．`AVERAGE` は平均を求める関数である．

分散を求める

- 平均は上のやり方で求めておく
- B1 のセルに `=A1^2` と入力する
- B1 のセルをクリックして「コピー」する
- B2 から B100 までのセルを範囲指定して「貼り付け」する
- B101 のセルに `=SUM(B1:B100)/100` と入力する
- C101 のセルに次のように入力すれば，分散が得られる．
`=B101-A101^2`
- 標準偏差を求めるには，C102 のセルに，`=SQRT(C101)` と入力すればよい．

分散=2 乗の平均から平均の 2 乗であることを使っている (p.8 を参照)．

度数分布表を使う

今度は度数分布表から統計量を計算しよう．第 1 章に出てきた表 1.2(p.16) のデータを，次のように入力する．

- エディタを開く
- 数値データのうち，階級値と度数のペアを，カンマ', 'で区切って 1 行に入力する
- テキストファイルとして保存する
- ファイルの拡張子を `.csv` に変更したあと，ダブルクリックしてスプレッドシートに読み込ませる．これで右図のようにシートに入力されている．

	A	B	C
1	44	1	
2	46	3	
3	48	6	
4	50	9	
	...		
10	62	2	
11	64	2	

度数分布から平均を求める

- C1 のセルに `=A1*B1` と入力する

- C1 のセルをクリックして「コピー」する
- C2 から C11 までのセルを範囲指定して「貼り付け」する
- C12 のセルに次ように入力すれば、平均が得られる
 $=\text{SUM}(C1:C11)/\text{SUM}(B1:B11)$

度数分布から分散を求める

- 平均は上のやり方で求めておく
- D1 のセルに $=A1^2*B1$ と入力する
- D1 のセルをクリックして「コピー」する
- D2 から D11 までのセルを範囲指定して「貼り付け」する
- D12 のセルに次ように入力すれば、平均が得られる
 $=\text{SUM}(D1:D11)/\text{SUM}(B1:B11)-C12^2$
- 標準偏差を求めるには D13 のセルに $=\text{SQRT}(D12)$ と入力すればよい

表計算ソフトの「分散」は標本不偏分散

一般的なスプレッドシートのマニュアルには分散を計算するための関数として **VAR** が用意されている。これを使うと、たとえば A1 から A100 までのセルのデータの分散は $=\text{VAR}(A1:A100)$ で求められることになりそうだ。

ところが、実際にその通りに実行してみると、上で紹介したやり方で得られた値よりも微妙に大きな値が得られる。実はこの違いは、一般の表計算アプリケーションにおける「分散」の関数が標本不偏分散を与えるようになっており、また、「標準偏差」を表す関数 **STDEV** も、標本不偏標準偏差を与える仕様になっている。

このような事情は、専門の統計パッケージでもあることなので、くれぐれも、定義を知らないままに関数を使って、分散には区別があることを知らないで「分散を求めました」という過ちは避けるべきである。

D.3 本格的な統計計算には R がおすすめ

序文でも述べたように、本書は初等的な統計学について、なるべくごまかすことなく解説することを主眼にしている。これだけの内容を理解していれば、社会生活において出会うさまざまな確率的な事象、あるいは公表される各種の統計に対して十分な基礎になるはずだ。

しかしながら、研究や実務で実際に統計を駆使しようとする場合、統計処理は膨大な数値計算を必要とするものなので、コンピュータを使った統計処理が必須である。したがって、基礎の理解を前提として、統計処理のためのソフトウェアを活用することが、各分野での応用を試みる人には求められる。

そのさいに、前節で挙げた表計算アプリケーションを使うことは好ましくない。その理由としては、

- これらのアプリケーションは数値演算の精度が低く、統計処理における大きなデータを使う複雑な処理にはもともと適していない。
- Microsoft 社の Excel にはかなりのバグが存在することが知られている。これについては、群馬大学社会情報学部の青木繁伸氏のサイトに情報がまとめられている。
<http://aoki2.si.gunma-u.ac.jp/Hanasi/excel/>
- これらのアプリケーションが出力するグラフや画像はきわめて品質が低く、プレゼンテーションや印刷に用いるとかなり見劣りがする。特に出版用の図版にはまったく使い物にならないといってよい。

といったことが挙げられよう。最後の点は、本質とは関係ないと思われるかも知れないが、現代においては Beautiful Visualization という言葉が重視されるようになっており、決して軽視すべきではない。

そのような状況を踏まえた上で、統計の実務家にも大きく浸透してきた統計処理パッケージとして R がもっとも勧められる製品である。ここでは R を使った処理について字数を割く余裕はないが、簡単な特徴と情報のソースを示すことにしたい。

R の特徴

フリーソフトで無料で入手できる フリーソフトというのは、単に無料のソフトという意味ではなく、ソースが公開されて、自由に複製したり改変したりできるソフトウェアのことだ。

もともと S というベル研究所で開発された優れた、ただし有償の統計処理のシステムがあり、R はその機能をオープンソースで実装して生み出された。多数の専門

家がそのために協力して開発に当たり、現在でも進歩し続けている。

もちろん無料で公開されていて、だれでも自由にダウンロードして自分のパソコンにインストールすることができる。

多くのプラットフォームで利用可能 R は、Windows, MacOS, Linux など、一般に使われている多くの OS で使うことができる。

多種多様な統計手法に対応している R は現代において使われる多種多様な統計処理が可能であり、現代的な統計手法を含むあらゆる分野で利用されている。しかも専門領域によっては、さらに特化したパッケージが開発されて公開されていることも多い。

プログラミング言語であり、処理に汎用性がある R の多様な利用を可能にしている大きな要因は、それがプログラミング言語であるという点にある。そのため、きわめて柔軟に処理を組み立てることができる。一例をあげると、R では `var` が標本不偏分散を与える関数であり、デフォルトで標本分散を与える関数は備わっていないが、その場合にも自分で関数を定義することが簡単にできる。それらをまとめてパッケージ化することも可能である。

高品質のグラフィックス機能 R はデフォルトのままでも多彩で見栄えのするグラフィックス機能を持っていて、さまざまなグラフや絵を作成することができる。さらに `ggplot2` のような優れたグラフィックスを可能にしたパッケージも利用できる。

R に関する情報源

初心者が情報を得るためのウェブサイトと本を紹介しておく。R については、膨大なドキュメントや書物があるので、むしろどれを選べばよいのか迷うほどなので、まずは下の情報を参考にするとよいだろう。

The R Project for Statistical Computing 「R プロジェクト」つまり本家のサイト (英文)。

本体、および各種パッケージのダウンロードはここを利用する。

<http://www.r-project.org/>

The R Manuals 本家サイトで公開されているマニュアル群 (英語)。HTML による情報だけでなく PDF ファイルも提供されており、細かい目的ごとに分かれているので、英語が苦手であれば使いやすい。

<http://cran.r-project.org/manuals.html>

RjpWiki R に関して日本語で情報交換することを目的とした Wiki サイト。筑波大学大学院人間総合科学研究科の岡田昌史氏によって運営されている。

<http://www.okada.jp.org/RWiki/?RjpWiki>

R-Tips 舟尾暢男が作成したコンテンツを収録したサイト。農業・食品産業技術総合研究機構の竹澤邦夫氏によって管理されている。

<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>

「R による統計解析」 青木繁伸, オーム社 (2011)

この本はまとまりがよく, 最初から読むには適している。電子ブックも出版社サイトから購入可能である。

「The R Tips—データ解析環境 R の基本技・グラフィックス活用集 第二版」 舟尾暢男, オーム社 (2009)

上記の舟尾氏による R-Tips サイトのコンテンツを, 著者自身が大幅に加筆したもの。非常に分厚いが, その分きわめて丁寧に書かれていて読みやすい。電子ブックも出版社から購入可能なので, そちらのほうが使いやすいかもしれない。

R 初心者のための ABC A. ジュール, E. イエノウ, E. ミースターズ著, 石田基広, 石田和枝訳, 丸善 (2012)

R の初心者を対象にした講習会のテキストという位置づけの本。最初からていねいに追いながら課題を実行していけば, R 特有のデータ構造の取り扱いに習熟できるようになっている。セミナー形式の学習テキストとして推薦したい。

