

本書の学習内容

本書で学ぶ内容は、2つです。

コンピュータを使った統計解析

1つめは、コンピュータを使った計算です。前作の『基礎から学ぶ統計学』では、どの統計手法も、関数電卓で計算をしてもらいました。電卓での計算は、面倒です。しかし、統計学の最初の学習では、この面倒さが大切です。あえて電卓を使うことで、統計計算の仕組みを、身をもって実感できます。



画像提供 (左から) : シャープ株式会社, カシオ計算機株式会社, キヤノン株式会社

しかし、研究や調査では、こんなことは、絶対に行いません。統計解析専用に開発されたソフトウェアを使います。研究では、どんな些細な計算ミスも、許されません。プロ仕様のソフトウェアが必要です。

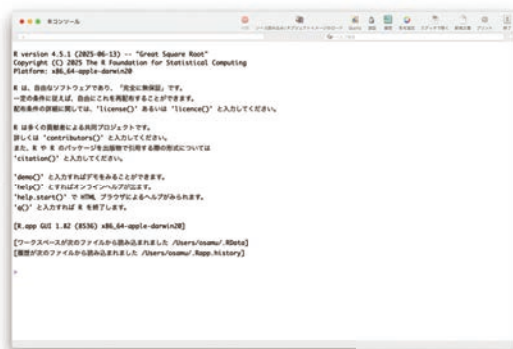
使うなら、国際的に評価が高く、著名なソフトウェアを使いたいです。私たちがレポートや学术论文を書くときには、行った統計手法と、計算を実行した統計解析ソフトウェアを(使用したバージョンまで)、しっかり明記します。計算手段には、胸を張りたいです。

この目的には、オープンソースでフリーウェアのRとjamoviがおすすめです。ともに無料です。この2つのどちらかで計算しておけば、たとえ、どんなに著名な国際誌に投稿する場合でも「統計計算はRで行った」とか「jamoviを使った」と、胸を張って書くことができます。

プログラム言語R

Rは、統計計算専用のプログラム言語です。本書の執筆時点で、世界中の多くの研究者が、Rで統計解析を行なっています。Rは、統計解析ソフトの世界

標準と言えます。



ただしRの学習は、プログラム言語の経験のない人には、敷居が高いです。しっかりと基礎を習得するには、おそらく、数ヶ月はかかります。ですが、「基本的な統計手法を実行するための、最低限だけを、身に付ける！」と目標を絞り込めば、学習が、かなり楽になります。例えば、 t 検定を例にします。表計算ソフトのExcelで計算に使うデータを作り、コピーしたら、あとはRで、例えば以下のように実行するだけです。

```
data <- read.delim("clipboard")
attach(data)
t.test(response ~ explanation,
        var.equal=T)
```

この程度の操作なら、覚える文法事項は、わずかです。第1章と第2章では、Rの最低限の初歩を、学びます。

統計解析ソフト jamovi

Rを簡単に操作するための試みは、これまで、多く行われてきました。そして、本書の執筆時点で、もっとも使いやすいのがjamoviです。jamoviは、完璧なGUI(グラフィカル・ユーザー・インターフェイス)を備えた統計解析ソフトです。あたかも、表計算ソフトのExcelを使うような感覚で、統計解析を実行できます。しかも、計算自体は、すべてRが行います。

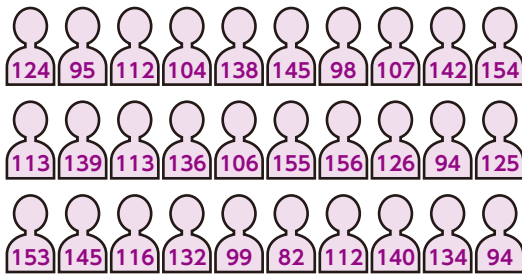


jamoviは、初心者にも使いやすいソフトウェアです。しかし、jamoviを学ぶ前に、Rの初歩の知識があった方がよいです。そこで、第1章と第2章でRを学んだあと、第3章で、jamoviの初歩を学びます。

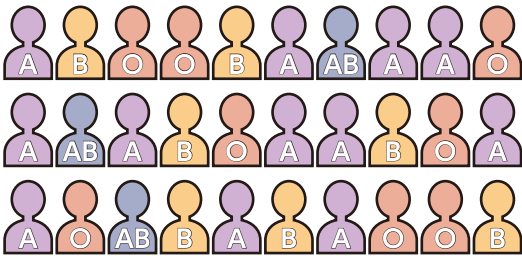
カテゴリカルデータの統計解析

本書の、もう1つの学習内容は、カテゴリカルデータの統計解析の基礎です。

前作の『基礎から学ぶ統計学』では、主に、観測値が「数値」であるタイプのデータを扱いました。例えば、コレステロール値の測定では、1人1人が、そのコレステロール値（数値型の変数）で特徴付けられます。



本書では、これとは異なるタイプのデータを扱います。例として、血液型の調査を考えます。この調査では、1人1人を特徴付けるのは、数値ではありません。

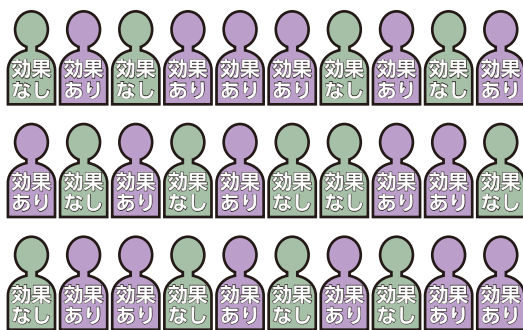


1人1人を特徴づけるのは、「A」か「B」か「AB」か「O」かの、4種類の区別です。このように、「種類の区別」を行う変数をカテゴリカル変数と呼びます。そして、カテゴリカル変数からなるデータをカテゴリカルデータと呼びます。

カテゴリカルデータは、様々なところで現れます。さらに2つの例を見てみます。1つめ。薬の効果を考えてみます。

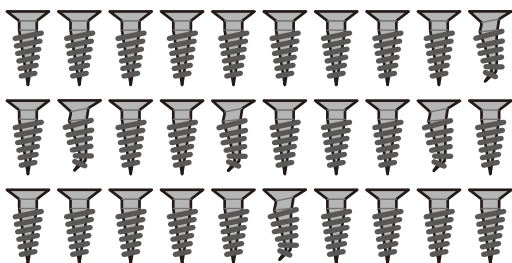


薬の効果を調べるために、患者の協力を得て、臨床試験を行ったとします。その結果

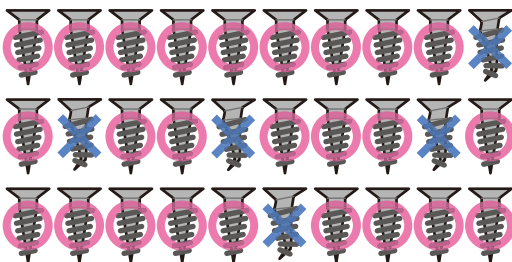


となったとします。この臨床試験で、各患者は、「効果あり」と「効果なし」の2種類を区別するカテゴリカル変数で特徴付けられます。

2つめの例です。工業製品を見てみます。工場で作られる製品は、一定の品質が保たれる必要があります。そこで、出来上がった製品は、1つ1つ、検査されます。

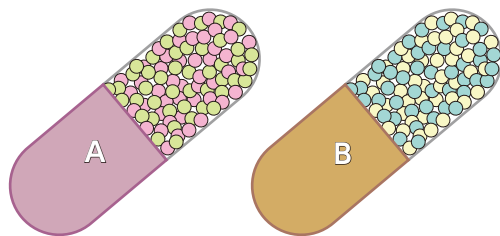


この検査で、製品は「良品」と「不良品」の2種類に区別されます。

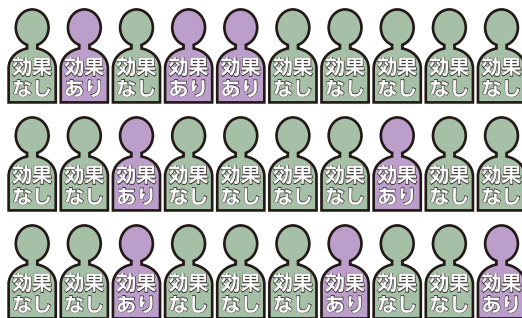


各製品を特徴付けるのは、「良品」と「不良品」の2種類を区別するカテゴリカル変数です。

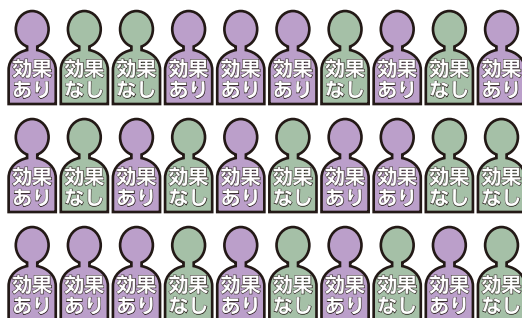
ここで、カテゴリカルデータの統計解析の、典型的な例題の1つを見てみます。例えば、2つの薬、A薬とB薬の効果の比較を考えます。



臨床試験を行ったとします。30人の患者がA薬を服用し、8人に効果がありました。



一方、B薬では、30人のうち17人に効果がありました。



この臨床試験での疑問は「A薬の効果と、B薬の効果は、異なるのか？」です。

本書では、第4章から第9章にかけて、カテゴリカルデータの基本的な統計手法を学びます。そして、第10章と第11章では、Rやjamoviで、カテゴリカルデータの統計解析を実行する方法を、学びます。