

まえがき



統計学に多変量解析と呼ばれている分野があります。変量とは、具体的には身長や体重など、測定される数値やカテゴリを意味します。ただし、体重は測ってみないことには正確なところは分かりませんし、また測るたびに違ってきます。そこで変わる数あるいは変わる量という意味で、統計学などでは変数とか変量と言います。

多変量解析とは、変数が複数集められているデータを解析する方法のことです。たとえば、学校では生徒ごとに複数の科目のテスト結果を集計しているはずですが、仮にその学校の1年生100人だけを対象として6教科のテストを実施したとします。すると、 $100 \times 6 = 600$ 個のテスト結果が集められるわけです。さて、読者の皆さんがこの学校の校長先生だとして、これら600個のテスト得点を眺めて、生徒たちの得手不得手を全体として把握できるでしょうか？ 600個のデータそのままでは、ここから生徒たちの学力に関わる情報を導き出すのはきわめて困難だと思われま

す。多数の変量から情報を抽出する手法が多変量解析です。もっとも多変量解析と一口に言っても、実際に応用されている手法は千差万別です。よく使われているのが「重回帰分析」や「クラスター分析」であり、さらに本書で紹介する「因子分析」、「主成分分析」なども有名です。

本書では主人公の高校生たちが、因子分析という手法を利用してクラス全体の学力を向上させようと画策します。その彼らの手助けをするのが、前著『とある弁当屋の統計技師（データサイエンティスト）』にも登場したデータ分析の専門家である二項文太と、彼がコンサルティングを引き受

けているお弁当屋さんの看板娘、正規乱子です。この計画を成功に導くため、高校生たちはデータ分析の初歩たる「平均値」の性質や2変数の「相関」について学び、さらには古典的な統計分析手法である「t検定」や「分散分析」についても解説を受けます。そして計画実現のかなめとなる多変量解析については、その「行列式」による表現を知り、そして因子分析の目的と解析結果の意味について知識を深めていきます。思惑通りにクラス成績を上げることができたかどうかは、本書の最後でご確認いただくとして、まずは彼ら高校生と一緒に多変量データ分析のなんたるかを体験してみてください。

なお本書の内容を補足するためのサイトを用意しています。サポートサイトでは、本書の範囲では説明しきれなかった細かな部分の解説を掲載しています。また本書の分析内容やグラフをパソコンで再現できるソフトウェアも公開されています。サポートサイトのURLと利用方法については、あとがきに付記しました。本書を一読された後、サポートサイトのほうにもアクセスいただけましたら嬉しく思います。