

はじめに

近年のグローバル社会では、多種多量のデータがかなりの費用と労力をかけて、一次情報として蓄積されている。これらのデータベースは、新しい観点からの発見的な現象情報に貴重な資源と考えられる。したがって、これらの重要な情報資源の分析には十分な配慮が望まれる。

とくに、情報化社会では、種々の構想のもと、分析に先立つデータの未知な数理的構造モデルの設定に困難が多く、とかく人為的になりがちである。また、新しい対象領域のデータでは要因変動と搅乱変動に関する分布法則の設定が、自然科学の物理や化学などと異なり、容易ではない。しかも、通常の情報資源のデータベースには、単に計数であったり、名目カテゴリであったり、水準や順序付け資料であったり、素朴な生の原始データには複雑で不確定な要素を多く含んでいる。したがって、このようなデータ群に、直ちに定型的な分析法の導入は誤った結果を導く怖れがある。このため、データ解析の初期には、分類表・クロス集計表・折れ線グラフなどが作成されるが、これらも単に感覚や印象に訴える恣意的な論議にすぎないことになる。このことは最近のデータマイニングの立場でも同様の論議となり、せっかく生の資料を観測・収集しても、新しい発見的な提唱には、取り扱いの上で科学的方法論による慎重な証明が必要になる。すなわち、収集したデータ群の搅乱変動を排し、提唱内容が真である推測法が求められる。

本書では、この種の主要なほとんどすべての計数的解法を網羅し、また各解法の解説では実践と学習に一々数値例を丁寧に付記し、手許で各解法の考

え方・解き方がごく自然に理解できるよう意図している。ちなみに、これらは数学的な仮定がほとんど不要な計数的理論に基づき、何の予備知識も要せず広範に適用される解法を集録している。

また、これらの情報処理法の選択とそれらの実践には容易に利用できるソフトウェアシステムが必要になる。このため、著者らは上記のすべての解法に必要かつ有用なソフトウェアシステムを、パソコンでも身近に実践できるように、実用現場での数百余の蓄積されたデータ集合を想定し、さらに大量のデータにはクラスタ分析や計数データを探索的に採取するのに便利な逐次方式や主要な一対比較法の解説も加えて開発している。同時に、日常の統計解析に付随して必要な、主要な分布（上側確率 パーセンタイル）の数値表や各種乱数を得るソフトウェアをも加え、FADA システムを構築している。昨今の組織的に蓄積される情報資源・データファイルによる情報解析の端緒に、本書はいわゆる初期的選択としての解法群と関連の約 70 編を含む全ソフトウェア（FADA システム）を WEB で提供している。具体的に、第 7 章にシステム内容の全メニューと適用要領を示している。

現在、多くの企業や役所などの情報や統計の担当部課で情報分析に日常取り組んでおられる方々や、大学で広範な情報処理を学ぶ学生諸君に、即戦的学習と実践に役立つように意図している。本書によって、読者が探索や検証を追求する実践科学の現場で、また学習や演習の教科書として、有効かつ有用な成果を挙げられるよう期待している。

むすびに、本書の校正を担当された立教大学教授山口和範氏に感謝の念を付記し、また本書の刊行にお世話下さった取締役信澤孝一氏および編集課長横田穂波氏のご親切に心から御礼を申し上げる。

2006 年 7 月

淺野 長一郎
竹内 光悦