

## データ解析のすすめ方

新村 秀一 著 共立出版 1989年 定価2000円

統計的なデータ処理を行なうに当って、各種の統計パッケージの利用が日常化している現在、統計手法の入門書もこれらパッケージの存在を十分に考慮したものであってほしい。特に、具体的なデータ解析を前提とするユーザーにその道具としての統計手法を目的に解説するのであれば、手法の導出に伴う一般に難解で不馴染みな数理はとりあえずパッケージのコマンドに置き換えて、そのため、理論的に厳密な表現は多少犠牲にはなるが、一方でより本質的な部分、すなわち、現象モデルと解析目標の設定、目的に即した手法の選択、データ形式の適用可能性のチェック、解析データの解釈等に重点を置くことができる。

本書は、汎用統計パッケージとして国際的にも広く普及している SAS (Statistical Analysis System) について、わが国で当初より関係し最も長い経験を持つ著者が、SAS の使用環境下で統計的データ解析を懇切かつ平易に解説した好書である。特に、適用例として共通した1つのデータを全編にわたり使用し、そのデータのコンピュータ入力から単純集計、そして、各仮説検証のための種々の統計手法の適用と、全体を現実のデータ解析に直面する一連の流れに乗せて解説することによって、類書に例を見ない特色を示している。多くの従来の入門的類書は、単に手法をモザイック的に配置しデータ解析の推理 (!) 過程を解説しないことで、読者に通し読みをいささか退屈なものにしているが、本書はこれを排除して活きた手法の利用法を提示している。また、用いたデータが一般に記憶に新しいつづきの科学万博全会期中の入場者状況を示した生データであることが、人為データの使用例では望めない各解析結果の真の迫力として、読者を楽しませてくれる。本書の通読によって、読者は、SAS と統計手法の各々の解説と相互関係を学びながらデータ解析の入門コースを同時に実感することができ、おまけに、解析を通して抽出された科学万博の入場者状況に関するいくばくかの情報を手に入れることができるのである。(Oh! My God!)

本書の構成を以下に示す；

### 1. 前準備

#### 1.1 データ収集からコンピュータ入力まで 1.2 SAS

#### S のプログラムとデータの修正

#### 2. データの出力

##### 2.1 PRINT プロセジャー 2.2 時系列プロット

#### 2.3 単純集計

#### 3. 数値データ分布の特性値とグラフ表現

##### 3.1 要約統計量 (MEANS) 3.2 詳細な要約統計量 (UNIVARIATE) とグラフ表現 3.3 層別箱ヒゲ図 (SPLOT) 3.4 ヒストグラム 3.5 処理の繰り返し

#### 4. 2変数の関係を調べる

##### 4.1 2重クロス集計 4.2 3重クロス集計 4.3 散布図 4.4 相関分析

#### 5. 回帰分析

##### 5.1 単回帰分析 5.2 日本の自家用車は8人乗り？

##### 5.3 分散分析 5.4 予測と回帰分析 5.5 分散分析と数量化 I 類 5.6 共分散分析

#### 6. 主成分分析

##### 6.1 相関行列 6.2 主成分の出力 6.3 主成分上でのプロット

以上の目次にみられるように、データ解析で日常的に用いられる手法はかなりの水準で本書に盛り込まれている。むしろ、盛り込みすぎといえるくらいである。後半、著者は、さまざまな統計手法の紹介を意図するあまり、本来の科学万博のデータ解析の流れからはやや不自然と思われる仮説を採用して解析の流れを混乱させている嫌がある。したがって、科学万博のデータ解析の専門書として本書をみる向きにはやや詰め甘さが気になるが、著者が最初から意図するように、社会人、大学初年次の大学生を対象とした SAS によるデータ解析の自習書として、本書はその役割を十二分に果たしている。

巻末の付録に、実際に用いた生データと SAS プログラムの全リストが掲載されている。読者は、手近かなコンピュータを利用して実際に本文の解析を再現実行することも、また、新たに自分なりの解析に腕を振るうこともできる。データ解析は解析者の個人芸である。本書はその個人技を磨く好材料となろう。

(渡辺美智子 関西大学)