

序 文

厚生省の統計によれば、中国地方の平均寿命は長い。なぜだろうか。

平均寿命には、気温、食べ物、地理条件、文化水準、医療状態など多くの要因が複雑な形でかかわりを持っており、先の問に答えるためにはその因果関係の解明をしなければならない。肺ガンや胃ガンの発病、あるいは最近の就学生の学力のつまづき、非行のひん発と言ったこと、さらには工場で生産される製品の欠陥の背景にもいろいろの要因がからまっていると考えられる。

多変量統計解析法 (methods of multivariate statistical analysis) はこうしたいいくつかの項目の間の関連性を統計的に分析し、現象を要約して簡潔な表現を与えたり、現象の背後にひそむ構造を浮き彫りにしたり、あるいはある項目を他のいろいろの要因から予測（説明）したりする方法で、上に述べたような複雑な現象を解明するための有力な方法論としていろいろの分野で巾広く応用されている。このような普及は近年パーソナルコンピュータや大型コンピュータの端末器などの計算手段が容易に利用できるようになったこととも大いに関係があると思われる。たいていの大型コンピュータにはアプリケーションプログラムとして一通りのプログラムが用意されているし、パーソナルコンピュータ用の BASIC プログラムもいくつか公表されているので、現在ではその気にさえなれば、誰でも多変量統計解析法を使うことができる。すなわち、データを準備して所定の様式でコンピュータに入力すれば、面倒な計算をしなくても解析結果を手に入れることができる。したがって、人はただ解析法の考え方を理解することと、コンピュータの出力を正しく読むことができればよい。こうした時代の中で多変量統計解析法は、文科系、理科系を問わず、データにもとづいて実証的に積みあげていくような分野においては、誰もが勉強しておくべき基礎的な素養に近くなってきているように思われる。

本書は著者らの行ってきた講義や各種の講習会での講演、いろいろの分野の研究者や技術者の方からのデータ解析に関する相談などの経験をふまえて、統計学や数学の専門家以外の一般の人にもできるだけわかりやすく、しかも重要なことはきっちり解説しようという方針で書いた物で、次のような点に留意している。

(1) 文科系の学生でも理解しやすいようにベクトルや行列による表現はで

きるだけ使わずに、初等的に記述した。

- (2) 重要な結果は天下り式の記述は避け、導出の筋道をていねいに述べた。
ただし本文中で述べると冗長になるところは演習問題にまわした。
- (3) 各章ごとに2種類の演習問題をつけ、巻末に詳細な解答を付した。問題Aは理論的な問題、問題Bは実際にデータを扱う問題である。
- (4) 各方法のねらいと手順を例示するため、数値例、応用例を豊富につけた。(統計学の理解のためには実際にデータを解析してみるのがよい。電卓やポケコン、マイコンなど手近な計算手段を用いて、例題や問題Bの計算をやってみられることをおすすめする。)

多変量統計解析法と一口に言っても、回帰分析、判別分析、主成分分析、因子分析、グラフ解析、クラスター分析、林の数量化法、…などいろいろの方法が含まれており、応用する人は解析の目的やデータの性質によってどのように使いわけるか(さらには使いこなすか)に困難を感じられるかも知れない。そのような際の1つのガイドラインとして、各種の方法の選択の手引きを次ページに示したので、活用されたい。教科書として利用していただく場合、半年の講義ではやや量が多いと思われるが、その時には回帰分析、主成分分析、判別分析、数量化I、II類などの基礎的な章を中心に、対象に合わせて取捨選択していただきたい。

本書は1980年12月～1982年2月の「BASIC 数学」誌に連載したものにもとづき、講義での反応などを考慮して大巾に加筆、修正したものである。その間数値例の計算や校正などでお世話になった田中潔、瀬尾日出夫、林篤裕、河田貞江(いずれも岡山大学学生)の各氏に感謝の意を表します。また連載中からいろいろご迷惑をおかけし、数値例のチェックまでしていただいた現代数学社の古宮修氏にあつくお礼申し上げたい。

1983年1月

田 中 豊
脇 本 和 昌

目 次

序 文

多変量統計解析法の選択の手引き

第 1 章 回帰分析法..... 1

§ 1	線形回帰（直線回帰）	2
§ 2	回帰直線による予測誤差の標準偏差.....	5
§ 3	相関係数	6
§ 4	変数変換と回帰	9
§ 5	線形重回帰	11
§ 6	重回帰式による予測誤差の標準偏差	17
§ 7	重相関係数	17
§ 8	偏相関係数	21
§ 9	回帰係数の推定と検定（単回帰の場合）	22
§ 10	回帰式による予測値の区間推定（単回帰の場合）	29
§ 11	回帰係数の推定と検定（重回帰の場合）	31
§ 12	回帰式による予測値の区間推定（重回帰の場合）	39
§ 13	説明変数の選択	42

第 2 章 主成分分析法.....53

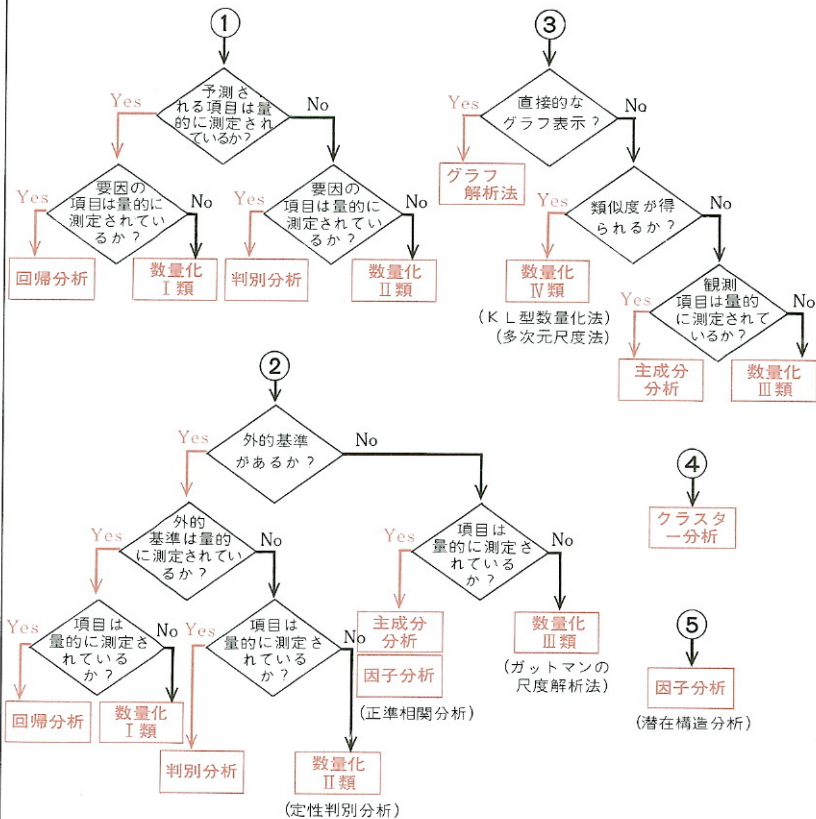
§ 1	主成分分析とは－2 変量の場合	54
§ 2	2 変量の場合の主成分の導出(1) －合成変量の分散の最大化.....	56
§ 3	2 変量の場合の主成分の導出(2) －もとの変量との（重）相関係数の 2 乗和の最大化	66
§ 4	変量の標準化	68
§ 5	p 変量の場合の主成分の導出(1) －合成変量の分散の最大化.....	71
§ 6	p 変量の場合の主成分の導出(2) －もとの変量との（重）相関係数の 2 乗和の最大化	77
§ 7	とりあげる主成分の数	80
§ 8	固有値問題の数値解法	81

§ 9 適用例	88
第 3 章 判別分析法	101
§ 1 判別分析の考え方	102
§ 2 1 変量による判別分析	102
§ 3 2 変量による判別分析	109
§ 4 k 変量による判別分析	121
§ 5 グラフによる判別方式	128
第 4 章 数量化法	137
§ 1 数量化の考え方	138
§ 2 数量化 I 類	139
§ 3 数量化 II 類	151
§ 4 数量化 III 類	161
§ 5 数量化 IV 類	171
第 5 章 因子分析法	179
§ 1 因子分析とは	180
§ 2 因子数のきめ方	185
§ 3 因子負荷量の推定	185
§ 4 因子軸の回転と因子の解釈	190
§ 5 因子得点の推定	195
§ 6 例題－食品嗜好データの因子分析	196
第 6 章 グラフ解析法	201
§ 1 はじめに	202
§ 2 直接多変量を平面上に記述する方法	203
§ 3 多変量を変換してから平面上に記述する方法	219
第 7 章 クラスター分析法	229
§ 1 クラスター分析とは	230
§ 2 クラスター分析の考え方	230
§ 3 クラスター分析の方法	235
問題の解答	245
参考文献	283
さくいん	289
数表	290

多変量統計解析法の選択の手引き

解析の目的は次のうちどれか？

- ① いろいろの要因によってある項目を予測（説明）したい。
- ② 観測されている複数の項目を代表する総合的指標を求めたい。
- ③ ものや項目の間の関係を視覚的にとらえ、それにもとづいて分類をしたい。
- ④ ものや項目を似たもの同志をまとめるように分類したい。
- ⑤ 項目間の複雑な相関関係を説明する潜在的構造を知りたい。



(注) () 内の方法は本書では省略

〒606 京都市左京区鹿ヶ谷西寺ノ前町1 振替京都1-11144

現代数学社