

残すことができる。

- (5) (3) で述べたことと一部重複するが、市販統計ソフトウェア、所内開発ソフトウェアの中から、必要なものを自由に選択し、組み合わせて使用することができる。

集団健診における循環器疾患の健康予測

駒 澤 勉

一般に、医師はその道の専門家としての経験にもとづいて、受診者の皮膚の色つや、問診などから主観的に判断したデータと客観的にいろいろな生体に関する特性を計測したデータを総合して診断を下す。この診断過程は、得たデータにその医師なりの経験にもとづいた重み付けを行い、それらを総合した重み付け得点—その医師なりの頭の中に描いた健康指標—を使って、あらかじめその医師が長年の間にデータ整理しておいた疾病群のどの類似群に属するかを判断している。名医ほどこの診断過程に使用する情報データの整理・分析法に優れているわけである。いいかえれば名医ほど数多くの受診者のさまざまな健康に関するデータを収集し、分類・整理し、いろいろな健康指標を作り出して実践的に活用している。時には、試行錯誤の繰り返しにより自分なりの健康指標を改良している。統計的方法のなかでも、多次元データ解析は、名医の診断過程に呼応するところが多い。

そこで、多次元データ解析の応用展開として、集団健康検診の検査データに適用して健康の指標化、予知、予防の実証的研究を行ってきた。研究は現在も継続中で、本年度は実践的な多次元データ解析の検討を行った。

- (1) 説明変量間に従属関係が成り立っていても利用可能な質的データの主成分的重判別分析(数量化第 III 類による判別分析)の研究。
- (2) 循環器系の集団健康検診の大量データによる診断応用の研究。

多変量解析の目的変量(外的基準)がある場合の方法論は、説明変量間には互いに独立であることが方法論上の条件である。そのため現象を予測や判別の多変量データ解析で分析する際、説明変量間に従属関係があると、必然的に説明変量間に従属関係のないモデル式を設定して解析することになる。しかし、医学分野に限らないが、特に医学分野の現象解析では従属関係の強い説明変量もモデル式に取り込んで解析する必要がほとんどである。そのため、要求に応えるべき予測・判別のデータ解析の研究を行っている。

領域統計研究系

社会移動の測定モデル

領域統計研究系(客員) 大阪大学人間科学部 直 井 優

社会移動表とは、主に父と子の世代間の職業的地位を交叉させた正方のクロス表をさす。形式的には、クロス表にすぎないから、一般のクロス表分析が可能である。しかし、社会移動研究において、この 40 年間、解析法の開発に大きな努力が払われてきた。それは、社会移動表から、子が父の職業的地位とは関係なく、職業的地位を達成できる度合、すなわち、機会均等の度合を測定するためである。このためにまず必要なことは、「完全な機会均等の状態=完全移動の状態」を定義することである。このために統計的独立性のモデルが用いられる。ところが、クロス表の周辺度数を用いた統計的独立性のモデルでは、父と子の周辺度数が異なる場合、その指数が大きく影響されることがわかった。そこで安田指数では、父と子の同一職業における周辺度数の差の絶対値(全体の場合にはその総和)をもって強制移動(または構造移動)と定義し、全体の移動量から強制移動量を差し引いた移動を純粋移動(または周流移動)とよび、これによって実際の社会移動の状態と完全移動の状態とを比較する開放性係数を提唱した。この係数は高く評価され、広く用いられている。ところが、この係数も、純粋移動の量をだすために、周辺

度数を用いているため、父と子の周辺度数の違いによって影響される。

そこで、本研究では、周辺度数が父と子の世代間で同一の場合に起因する移動を純粹移動と定義し、線形計画法を適用して、この純粹移動の量を求めた。解法には、シンプレックス法を用いた。これにより修正開放性係数では、社会移動の機会、実際の純粹移動の完全な純粹移動に対する比として求められる。この係数は、統計的検定が可能であり、趨勢分析や国際比較分析にも適用することができる。

3月24日

領域統計研究系

疎なデータの理論 — 展望

柳本武美

1. 問題

母集団が K 個の層に分れていて、標本が第 k 層から大きさ $n_k, x_{k1}, \dots, x_{kn_k}$ が得られるとする。第 k 層での密度（確率）関数が

$$p(\mathbf{x}_k; \theta, \mu_k), \quad \mathbf{x}_k = (x_{k1}, \dots, x_{kn_k})$$

であると仮定する。データが疎であるとは各 k について n_k が大きくない場合、例えば 20 以下を指す。従来、疎なデータについては K が大きい場合に関心もたれていた。無限に母数をもつ場合の理論がその 1 例であるが、理論的にはとにかく実際のでない。

2. 研究の必要性

データの質に関してみれば、実験技術と条件の向上によって、近年著しく向上してきた。これは社会調査とは事情を異にしている。データの質の向上は当然モデルの精密化を要請する。従来、誤差として扱っていた要因の違いを、要因による層別として処理する方が良くなる。

3. 研究の方法

近年の計算技術の向上と計算機器の著しい向上によって、複雑な計算が廉価にかつ容易に遂行されるようになった。また上に述べた研究の必要性に刺激されて、統計解析の理論の進歩が促された。1つの極めて大きな進歩は修正尤度法の利用である。

- イ) 経験ベイズ法: 多くの母数 μ_k を推定するために、 μ_1, \dots, μ_k を超母集団からの大きさ K の標本とみなして、尤度法を 2 段階に適用する。この方法の有効性は良く知られたスタイン問題にその嚆矢をみるが、病理的な手法ではなく適用範囲の広い実用的な手法を生みだしている。
- ロ) 条件付尤度法: 母数 θ を μ_k の影響を小さくして推定する規準として、条件付最尤法が利用される。条件付尤度が利用できるようにモデルに制約があるが、実際に多くの有用なモデルをカバーする。

その他、モーメント法が頑健であることが分り、従来の大標本理論で説明できない結果が得られる。