

ベクトル積率とその統計量

田口時夫

K. ピアソンによって確立された古典的なモメント概念を基本とする統計記述や解析の方法は、数理統計学に、初めて数学的に整理された体系を与えることになった。これに対して、例えばフラクタルモメントやノルム統計量のような構想は、従来から存在していたが、概して断片的な成果に止まり、ピアソン系に比肩するような統一の体系を形成しているとはいえない。その間であって、ローレンツ曲線を対象とする幾何学的分析に基づいて形成されたジーニの集中統計量や集中解析の方法は、その当初から一部の研究者には、大きな可能性をもつものと予想されてきた。ただし、その体系化を実現する為には、それを多変量解析に導く、集中曲面・超曲面の具体的な規定が必要であったといえる。

筆者は従来から、この曲面の妥当な規定と解析を試みてきたが、その研究成果に基づいて、本年度はある程度、体系化の実現を達成し得たと考える。具体的に統計集団の個体間に符号関数やベクトル積演算形式を導入して系統的に形成される平均量によってベクトル積率を規定し、更にこれらの積率間の単純な演算によって得られる統計量を用いて、一つの包括的な記述系を構成することができた。これらの統計量の一部は、従来の集中解析においてある種の統計指標として既に適用されているので、これらを総括して、集中統計量とすることができる。これらの系は、ピアソン系に比較すると、一般に広い適用範囲とゆるい条件をもつことが見出される。集中統計量は、更に平均差を中心とする線形構造の記述・解析系と、ジーニ係数を中心とした非線形構造の記述・解析系に分離される。両者を区別する性格の一つは、前者が測定値を直接演算対象とするのに対して、後者は測定値をパーセンテージで表示した結果に同じ演算形式を適用する点にある。以上の諸結果から、更に一般的な積率や統計量の展望を得ることができた (Taguchi (印刷中))。

参 考 文 献

- 田口時夫 (1988). 集中多様体と集中解析のシステム (I) —— ジーニの統計方法論の幾何学的展開 ——, 統計数理, 36, 41-52.
- 田口時夫 (1988). 弾力性係数をめぐる疎な理論, シンポジウム「数理統計学と計算機の接点に関する研究」報告集, 科学研究費総合(A)「統計学の数理的基礎と諸分野への応用に関する研究」, 研究代表者: 広津千尋.
- 田口時夫 (1989). 集中多様体と集中解析のシステム (II) —— 相対的ベクトル積率と非線形集中統計量 ——, 統計数理, 36, 140-166.
- Taguchi, T. Vector product moments and system of statistics (ed. Y. Dodge), *Statistical Data Analysis and Inference* (in print).

統計データ解析センター

数量化第 III 類による判別空間の選択

駒澤 勉

多変量データ解析を利用して、実際の現象を分析する際、特に重回帰分析や、重判別分析のように外的基準がある分析において、説明変数に従属性の強い変数が含まれている場合には、分析結果の要因的意味解釈に問題が生じることが多い。外的基準のある多変量データ解析法は、方法論からいって、次式の線形モデル式の説明変数間に (X_j, X_u) が独立でなければならない。

$$Y = \omega_1 X_1 + \omega_2 X_2 + \dots + \omega_m X_m.$$

しかし、特に医学現象解析においては説明変量間に従属関係があっても、判別・予測問題に、臨床的立場から必要な変量が他の説明変量と独立でない場合でも線形モデル式の要因として選択しなければならないことが非常に多い。このような場合の解決策として、主成分分析の情報を利用して判別分析を行ない、問題を解決することがある。ただし、共通の分散共分散行列（または、相関行列）に対して主成分分析を施す際に、群間の変動と群内の変動の方向によっては、分散 σ^2 （寄与率）の必ずしも大きい主成分 Y を用いれば良いとは限らないが、重み ω_j , $j=1, 2, \dots, m$ （説明変量 X_j と主成分 Y との相関係数 $r_{X_j Y}$ ）が現象の判別に関与している主成分軸であることを現象の専門的知識から選別することで、判別・予測問題に対して多変量データ解析が成功することが非常に多い。

このことは、質的データの多変量的解析の数量化の方法論でも同様な問題を生じる。数量化の場合、説明アイテム間の従属関係が強いアイテムを次式の線形モデル式へ選択すると、

$$Y = \sum_{j=1}^m \sum_{k=1}^{l_j} \omega_{(jk)} \delta_{(jk)}.$$

外的基準の数量を予測する第 I 類、外的基準の項目・区分を判別・予測する第 II 類では、強い従属関係の説明アイテム・カテゴリー ($C_{(j1)}C_{(j2)} \dots C_{(jl_j)}$), ($C_{(u1)}C_{(u2)} \dots C_{(ul_u)}$) に与える数量 ($\omega_{(j1)}\omega_{(j2)} \dots \omega_{(jl_j)}$), ($\omega_{(u1)}\omega_{(u2)} \dots \omega_{(ul_u)}$) が要因的解釈において正反対の符号を持った数量となる。これを解決するには、主成分判別分析と同様に、数量化第 III 類の解析解を利用して分析する。最適な成分軸は試みの基準での重相関、相関比、判別の中率、分割表の AIC 等の評価値、および先験的情報で選択する。この実証的研究として、循環器系の健康診断の事例を示した。

「日本人の国民性調査」のサンプリング計画

中 村 隆

1988 年秋に実施された「第 8 回 日本人の国民性調査」のサンプリング計画について、パーソナルコンピュータ上のデータベースシステムで作成したサンプリング計画用プログラム SAMPLAN/PC を操作しながら報告を行なった。

「日本人の国民性調査」は、昭和 28 年から 5 年ごとに統計数理研究所が実施している全国規模の継続調査である（母集団は 20 歳以上の有権者）。計画標本の大きさは 6,000 であり、これを折半して K 調査票（継続質問中心）と M 調査票（新しい質問中心）に割り当てて訪問面接式調査を行なっている。調査対象者のサンプリング方法は、層別多段抽出法であり、ほぼ同一の方法を現在まで続けている。

サンプリング計画は、全国 3,300 余りの市区町村の有権者数と層別に必要な情報の収集から始まる（この段階で、毎日新聞社世論調査部のお世話になった）。今回は、ラップトップコンピュータを持ち込み、SAMPLAN/PC によって市区町村データベースを構築した。このような方法をとることによってデータの入力や整合性のチェックが柔軟に行なえた。

次の段階は、市区町村の層別である。従来は県別を主要な基準として区部(6 層)、市部(29 層)、郡部(20 層)、沖縄(1 層)の計 56 層に分けていたが、地点間での標本の大きさのばらつきが大きくなるので、今回は層を区部、人口 20 万人以上の市部、人口 20 万人未満の市部、郡部、沖縄の 5 層に簡略化した。ただし、従来とも各層の中ではさらに県別・人口規模別に市区町村を並べ換えている。層の組替えなどの検討も SAMPLAN/PC を使うことによって自由に行なえた。

層別が終了すると、実際に第 1 次抽出単位である地点(市区町村)を抽出する。層ごとに物理乱数を用いてスタート番号を決め、市区町村を等間隔で抽出していく。SAMPLAN/PC によって、抽出された市区町村名、有権者数、スタート番号、割当サンプル数などの一覧が得られる。前回は大型計算機によって同様の抽出を行なったが、今回は漢字で市区町村名が得られるという利点が大きかった。また、サンプリング計画の細部での変更に対しても迅速に対応ができた。