

# データ解析法

東邦学園大学経営学部 田浦 元

## 報告概要

近年、マイクロ統計分野の研究が活発に行なわれ、統計解析の手法においても、マイクロデータを利用した分析が多く試みられている。記念号第3集における統計解析の論点の中心は、計量モデル分析や時系列分析等であったが、これらは集計量に基づく分析手法（あるいは集計量に基づく分析を基礎として発展してきた分析手法）といえよう。近年、これら集計量に基づく分析の限界が見出される中で、マイクロデータに基づく分析の必要性が高まり、多くの分析が行われている。本報告ではこのマイクロデータに基づく分析に焦点を絞り、特にマイクロデータを利用したパネルデータ分析や質的選択モデル分析の現状についてみることにする。

また近年、データマイニングの手法についても多くの研究が行なわれている。データマイニングも大量データの利用により実現した手法であり、その点においてマイクロデータに基づく分析と共通する背景を持つ。そこで、データマイニングの手法についても概観する。以上を通じて、これらのデータ解析法の意義と限界について考えることとしたい。

## 1. マイクロデータに基づく分析の活発化

- ・ 集計量に基づく分析手法（およびそれを基礎として発展した分析手法）
  - ・ マクロ計量モデル分析
  - ・ 時系列分析
    - ・ VAR（多変量自己回帰：Vector Auto Regression）モデル（Sims1972,1980a,b）
    - ・ ARCH（自己回帰条件付分散不均一：Autoregressive Conditional Heteroscedastic）モデル（Engle1982）→金融時系列モデル分析
- ・ マイクロデータに基づく分析手法（およびマイクロデータの利用により活発化した分析手法）
  - ・ パネルデータ分析（マイクロ）
  - ・ 質的選択モデル分析
  - ・ (データマイニングの手法)

## 2. マイクロデータ分析とデータ解析

### (1) パネルデータ分析

- ・ クロスセクションデータを時間的に継続して収集。時系列データにクロスセクションデータの特徴を持たせた分析が可能。
- ・ プールモデル、固定効果モデル、変量効果モデル
- ・ 課題：サンプルの脱落、サンプルの不追加
  - ・ 脱落の補完についての研究（Little-Rubin1987等）

## (2) 質的選択モデル

- ・被説明変数が非連続な場合に、連続的な潜在変数 (latent variable) を想定し分析を行なう手法。
- ・ロジット (Logit) モデル (McFadden1974 等)
- ・プロビット (Probit) モデル
- ・トービット (Tobit) モデル (Tobin1958)
- ・二項選択 (Binary choice) →多項選択 (multinomial choice)
- ・順序反応 (Ordered Logit/Probit) モデル
- ・ネステッド・ロジット (Nested Logit) モデル (McFadden1983)
- ・質的選択パネル分析 (パネル・ロジット、パネル・プロビット等)
  - ・課題：誤差項の固定効果の問題

## 3. データマイニング

### (1) データの蓄積とデータマイニング

- ・基幹系データベースにおける大量データの蓄積
- ・データウェアハウス (Data Warehouse) の構築 (Inmon1996)
- ・データマイニング (Data Mining) : 蓄積された大量データから、データのパターンを見つけ出し、一定の規則を導き出そうとするもの。人間には思い付き難い出来事の発見。KDD (Knowledge Discovery in Databases)。

### (2) データマイニングと統計学的分析

- ・データマイニングの方法：
  - ・相関分析、回帰分析、クラスタ分析、因子分析等、統計学においても良く用いられる方法が中心的。
  - ・決定木 (Decision Tree) (Morgan・Sonquist1963)
  - ・ニューラルネットワーク (Neural Network) の応用
- ・抽出：
  - ・統計学的分析 : 標本調査 (母集団→無作為抽出→標本)
  - ・データマイニング : 既に蓄積されている大量データが分析対象
- ・プロセス：
  - ・統計学的分析 : 仮説の立案→モデルビルディング→データの取得 (標本調査) →データの解析 (集計、推定、検定) →結果の解釈・評価 (仮説の検討)
  - ・データマイニング : データの取得 (データウェアハウス) →データの解析 (特に探索型 DM) →結果の解釈・評価→仮説の立案