

## 論文の内容の要旨

論文題目 データ解析における本質抽出のための実践的アプローチ

氏 名 鈴木 知道

データ解析は、測定されたデータから、得られているデータの背後にある対象について理解することである。データ解析において用いられる統計手法の理論的・体系的な研究は数多く行われており、優れた教科書も多数出版されている。しかしながら、現実の問題を解決するにあたって、必ずしもこれらの理論を常に効果的に用いることができるわけではない。統計手法そのものは、問題解決の指針を直接は与えてはくれない。統計手法の適用が可能のように、問題を適切に統計的データ解析の言葉にブレークダウンする必要がある。すなわち、統計的データ解析手法において理論と実践のギャップが存在するといえる。しかしながら、与えられた問題に対して、問題をどのように統計的データ解析の言葉にブレークダウンするか、そしてどんなモデルを仮定するのか、どのようにデータを得るのか、解析対象変数はどのように決定するか、どの手法を適用すべきか、などに関して指針は明確ではない。

本論文では、データ解析を問題設定、解析方針決定、解析、結果の解釈というフェーズに区分し、自らが実施した統計的データ解析を通じて、データ解析手法を用いるにあたって何が足りないか、どこにギャップが存在しているのかを調査検討した。本論文は、その結果得られた、問題を解決するための基本的な考え方、必要な視点、方法論、知見をまとめたものであり、統計的データ解析における理論と実践のギャップに対する橋渡しを目指したものである。

第1章では本研究の背景と目的を述べている。データ解析とは、測定されたデータから、得られているデータの背後にある対象について理解することであり、解析対象からその本質的情報を導くことを目的としている。データ解析においては、「問題設定」、「解析方針決定」、「解析」、「結果の解釈」というフェーズを経るものとしてモデル化して、各フェーズにおける実践上の難しさを論じ、本研究の課題を明確にしている。

第2章ではデータ解析における現状の把握と問題点の抽出を行った。現実のダイナミックプロセス制御プロセスの調査から、データ解析のフェーズの要因を考慮し、問題点を抽出しまとめた。

第3章では、体系的な研究が行われているシステム同定の理論が、実践の場で有効に活用されていないというギャップが存在するという現状を調査し、同定に用いるための質の良いデータを得るための方法論が必要であることを明らかにした。そして、現実のプロセスにおいて設定しうる実験条件の検討を行った。考えられる多くの要因に対して、取り上げるべき要因を適切に絞り込み、制御因子としては2つのパターンを検討すればよいことを示した。そして、システムの未知パラメータの推定精度を最も良くする実験パターンを提示した。この解析を通して、実験が行えるというアプローチがとれるかどうか、そして、可能な場合にどのように実験を計画するべきかに関しての指針を明確にした。

第4章では、通常の時系列分析のフィードバック制御のモデルにおいて仮定されているが現実の適用の場では崩れている場合があることを述べ、工程平均を変化させる外乱が存在するプロセスがあることを示し、そのモデルを与えた。そして工程平均が変化する場合の制御式と、制御量を計算するための手順を提案した。制御式に関しては、厳密解と簡易解の両方を示した。厳密解に関しては、厳密な制御式に求め、その制御量を求めるためのアルゴリズムを提案した。また簡易解に関しては、得られた制御式に対しては割引係数を導入することにより目的特性値の平均二乗誤差を最小にする実践的な適用方法を提案した。

仮定したモデルが妥当であるかを診断するにあたって、残差分析を行う上での着目すべき視点の提案を行った。また、解析目的に適したモデルの立て方に関する指針を明らかにした。

第5章では、精密機械組立工程における時系列手法を応用した最適制御方式を求めた。通常、時系列分析が適用されない、非常に短時間で大量に製品が生産される組立工程においては、その1つ1つを時系列として扱うのは現実的ではなく、したがって大量のデータをどのように扱うかを考慮しなければならないことを明確にした。解析対象とする単位時間の大きさに着目し、データの信頼度と迅速性のトレードオフなどを考慮し、最適となる解析対象時間間隔を求めた。また、解析対象時間間隔とデッドタイムの関係も考慮し、

デッドタイムが整数で表せるような解析対象時間間隔がよいことも明らかにした。そしてこの解析を通じて、対象の変化が的確に記述できる大きさに関する知見を得た。

第6章では、中間特性値を導入した感性品質の評価および設計のための新しいモデルを提案した。研究対象としてとりあげた感性品質は、乗用車の運転席シートの座り心地である。この対象に対し、実際に中間特性値を構成すると考えられる物理量を実際に測定するという実験を計画し実施した。得られたデータを解析することで、着目すべき中間特性値とはいったいどのようなものであるかを明らかにすることが出来た。また、それらの中間特性値と感性品質評価値との関係などから座り心地とはどういったことであるのかに対して考察を行い、得られた結果に基づいてシート感性品質評価モデルを構築した。これにより、今後は中間特性の計測値を被験者による感性品質の評価に代わる新たな品質目標値とすることで、感性品質の設計段階における工数の削減、目的とする製品仕様の決定を効果的に支援することが出来る。そして、対象の理解のために、最終的な目的特性値に至る因果連鎖の関係を理解し、適切なレベルで中間特性値を設定することの意義とその方法に関する指針を与えた。

第7章では、東京都で発生しているヒートアイランド現象に対し、統計解析的アプローチをとることによって、東京都の気温が地域と時刻によってどのように変化するのか、その変化のしかたはどのような特徴を持つのか、を明らかにした。着目したのは、1日を単位とする気温の変化のしかたであり、これを気温日変化パターンとして抽出する。解析対象データは、東京都各所100地点において観測された1時間間隔の気温データを用いた。

真夏の気温データを用いた解析では、統計解析手法の一つである主成分分析をこのデータに適用した結果、気温日変化パターンの変動を代表する特性値である3つの主成分が得られた。3つの主成分の吟味から、これらは気温レベル、気温の日較差、気温変化の緩急と意味づけができた。そして観測された100地点に対し、主成分分析の解析結果をもとにグループ化を行った。その結果、観測地域をそれぞれ特徴を持つ7つのグループに分類できた。また、同時期のアメダスのデータを用いて解析の妥当性を検証した。

また、1年間のデータを用いて、1日の気温変化の特徴を表す気温日変化パターンが1年を通してどのように変化するのかを明らかにした。気温日変化パターンの抽出を、季節と天候の組合せに対し複数日を対象として行い、それぞれの日について最大3個の気温日変化パターンを抽出した。どんな気温日変化パターンが抽出されたかによって、対象とした日を吟味すると、それらは大きく以下に示す3つのグループに分類できることが分かった。1) 真夏の晴天日、2) 真夏以外の晴天日、3) 雨天・曇天日。そして、これらの解析結果をもとに、東京で起こっているヒートアイランド現象に関する考察を行った。

そして、これらの解析から、すでに得られているデータから、解析目的に合致する本質的情報を抽出するために、データの構造や性質に応じて目的に適した統計的モデルを構築する際の視点、指針を得た。

第3章から第7章での現実の統計的問題の解析の結果、それぞれ得られているデータから、対象に関して知りたい情報を抽出することができ、対象を適切に理解することができた。これは、データ解析における本質抽出のために、何が足りないのかと、何が必要なのかという立場から、さまざまな試行錯誤、工夫を行った結果得られたものである。これらの観点、アプローチがあったからこそ解析対象についての十分な情報が得られた。獲得できた工夫、コツを、データ解析における本質抽出のための教訓としてまとめた。得られた教訓は以下に示す。

#### A. 問題設定フェーズにおける教訓

- 教訓1 「データ取得のための実験を行うことが可能であるかを検討せよ」
- 教訓2 「個別最適化より汎用最適化が可能であるかを検討せよ」
- 教訓3 「中間特性値が設定可能であるかを検討せよ」
- 教訓4 「分割したデータの相違から全体のデータの挙動が把握可能か検討せよ」

#### B. 解析方針決定フェーズにおける教訓

- 教訓5 「モデルにおける誤差の仮定の確認を常に怠るな」
- 教訓6 「本質を失わない仮定を積極的に活用せよ」
- 教訓7 「変数の意味を吟味して、因子を明確に切り分けよ」
- 教訓8 「解析の範囲を明確にせよ」
- 教訓9 「対象の特徴・変化が的確に表現できる大きさを考慮せよ」

#### D. 解釈フェーズにおける教訓

- 教訓10 「自然法則、固有技術、過去の知識を積極的に活用せよ」

これらの教訓は「問題設定」、「解析方針」に関するものに集中した。これから、データ解析における本質抽出という視点から問題をとらえるとき、現実の問題をどのように統計的データ解析の言葉で表現し、そしてモデルを設定しデータを決定するか、という問題が重要であるということであることがあらためて浮き彫りとなった。これは、まさに統計手法をどのような場でどのように使うかに関する問題であり、統計手法の理論と実践のギャップが、ここに集約されているといっても過言ではないであろう。提案した教訓はまさにこれらのギャップに対しての解決案であり、これらの教訓の提案そのものがデータ解析における理論と実践のギャップの橋渡しとなり、実践的アプローチを示しているといえる。