

計算機統計学特論

(Advanced Computational Statistics)

種別・単位：講義・2単位（週1講時）

開講期：第2学期

担当者：水田 正弘（複合情報学専攻・大規模情報システム学講座・内線3755）

南 弘征（複合情報学専攻・大規模情報システム学講座・内線3756）

主題と目標

計算機統計学とは、計算機指向のデータ解析法であり、探索的データ解析やデータマイニングにおける基盤概念である。本講義では、計算機統計学の基礎的な知識、統計ソフトウェアの概説に加え、多次元データ解析法において計算機指向とされている手法を具体的に扱う。

想定する到達目標は以下の通りである。

- （1） 計算機統計学に関する基礎的な理解と応用可能性を知る。
- （2） 計算機を援用したデータの多角的解析の方法論を会得する。
- （3） 統計ソフトウェアについて正しい理解の下で適切に利用する背景知識を得る。

授業計画（項目、授業実施回数、内容）

項目	回	内容
計算機統計学とは	2	計算機統計学の基本的な概念を述べるとともに、統計パッケージについて説明する。また、統計パッケージRの基本的操作方法を学ぶ。
回帰分析	4	単回帰分析、重回帰分析について講述する。また、回帰分析における変数選択法、ロジスティック回帰分析を扱う。
判別分析とクラスター分析	3	データマイニングなどにおいて多用されるクラスター分析に関する理論と応用について講述する。また、それとの対比で判別分析について説明する。
主成分分析と射影追跡法	3	高次元データの次元縮小法として、主成分分析および射影追跡法について扱う。さらに、独立成分分析、射影追跡法の拡張手法である相対射影追跡法についてもふれる。
実践的事例概説と研究動向の紹介	3	計算機統計学の実例、最先端の研究動向などを講述する。実例としては、医学データ、経済データなどを教材とする。最近の研究動向として、関数データ解析法やシンボリックデータ解析法について扱う。

評価・教材・受講条件等

《評価》 講義に関するレポート・試験などにより総合的に学習の到達度を評価する。

《教材》 必要に応じて講義資料を配布する。また、適宜参考書を示す。

《受講条件等》

受講条件：数学および統計学の基礎的な知識を仮定する。

準備学習（予習・復習）等の内容と分量：予習と復習については、十分な時間をかけることが期待される。各授業の最後に、次回で扱うテーマについて簡単に述べるので、講義資料などを利用して、十分な準備をされたい。また、統計学および数学の知識が不十分な場合には、さらに準備が必要となる。講義資料にある課題は、次回までに解くことが求められる。

キーワード 判別分析、クラスター分析、回帰分析、主成分分析、射影追跡法、統計パッケージ