

統計科学と現象の分析 ビッグデータの活用法

ナンバリング 2 単位

DIF319



担当教員： 甫喜本 司

専 門 統計科学、データ科学、統計数理

出身校等 東京工業大学 大学院 理工学研究科 修士課程修了

現 職 北海道情報大学 情報メディア学部 教授

授業形態	前期インターネットメディア授業・後期インターネットメディア授業・後期 IP メディア授業
授業範囲	
メディア授業内容すべて	
試験期間	
シラバス検索画面トップページ (https://syllabus-tsushin.do-johodai.ac.jp/) 下部の「2025 学年暦」を参照	
試験範囲	
授業範囲すべて	
【IM 授業：試験時参考許可物】	
一切自由 ※ただし Web ページ（通信教育部 POLITE を除く）と生成系 AI の参照は不可とする。	
【IP メディア授業：試験時持ち込み許可物】	
自筆ノート（印刷物の添付不可）、電卓（機能制限なし）	
科目の概要	
現代では、現象の特性や現象に潜む様々なリスクを分析するために、ICT(情報通信技術)から出力される大規模なデジタル情報（「ビッグデータ」）を積極的に活用することが多くなりました。一見すると数値や文字の集合にしかみえないデータも、ある方法を実行することで我々に様々なメッセージを伝えてくれることがあります。本科目では、データから新たな知識を得るための方法を体系化したデータサイエンスの基本的な考え方について、計算機による実習を通して学びます。	
科目における学修の到達目標	
1. データの構造に応じて、適切な処理や分析を計算機で実践できる 2. データ分析の方法について、説明することができる 3. 現実の問題をデータに基づいて客観的に評価することができる	
講義の方針・計画	
この講義では、データ解析系言語である R を用いて、データから新しい知識を得る方法について学ぶことが中心となります。実際にコンピュータプログラミングを行いながら、データ分析の流れを学ぶことに加え、データを分析する方法、データ分析を通して目に見えない現象のメカニズムを評価する過程を学び、ノートを作成して下さい。試験ではノートの持ち込みを認めます。	
● 第 1 回：「ビッグデータ」時代におけるデータ科学	
大規模データを提供する代表的なサイトにはどのようなものがあり、どのようなデータが入手できるのか、そして我々にどのような恩恵をもたらすのかといった点を中心に体験をしながら、ICT(情報通信技術)が進む我々の社会における情報収集の現状を学びます。また、こうした大規模データの利活用を通してどのような発見ができるのか、本講義のガイダンスを兼ねて紹介します。	
● 第 2 回：R 言語入門	
本科目では、PC 上でデータ分析を行う際のプログラム環境として R 環境を使用します。R によるプログラミングにおいて、最小限知っておいてほしい基本構造について学習します。	
● 第 3 回：データの概観(1) – 1 次元データの可視化 –	
● 第 4 回：データの概観(2) – 多次元データの可視化 –	
● 第 5 回：確率的な分析 – 確率分布の推定と応用 –	
● 第 6 回：関係性のモデル化(1) – 散布図と相関 –	
● 第 7 回：関係性のモデル化(2) – 統計モデルとは何か –	

- 第8回：関係性のモデル化(3) – 統計モデルの推定と予測 –
- 第9回：関係性のモデル化(4) – 統計モデルの選択とAIC –
- 第10回：関係性のモデル化(5) – 非線形、および2値データのモデル –
- 第11回：時間変動の解析(1) – 時系列データの相関 –
- 第12回：時間変動の解析(2) – 周期の推定とスペクトル –
- 第13回：時間変動の解析(3) – 定常な時系列データのモデル化 –
- 第14回：時間変動の解析(4) – 非定常な時系列データのモデル化 –
- 第15回：時間変動の解析(5) – 周期性の強い非定常時系列データのモデル化 –

準備学習(予習・復習)

(予習) R言語に関する文献やウェブサイトは国内外に多数あります。各回のタイトルを基に、学習する内容について下調べして下さい。2時間

(復習) 授業後に、指定された「宿題」を各自のR環境で実習してください。また、学習したデータ解析の方法や考え方をノートにまとめて下さい。2時間

R言語はフリーソフトであり、自宅のパソコンでも作業環境を容易に準備して
学習することが可能です。関心のある受講者は積極的に取り組んで下さい。

成績評価の方法およびその基準

次項の項目および割合で標準評価基準に基づき総合評価する。科目試験による評価100%

課題(試験やレポート等)に対するフィードバックの方法

IM授業は、練習問題を解答すると、解答に応じたコメントを画面に表示する。

IPメディア授業は、授業中に課題を出す場合は、授業中に回答する。

教科書

書名：データサイエンス演習(改訂版)

著者名：甫喜本司

発行所：学術図書出版社

ISBN：978-4-7806-1243-1

参考書・参考Webサイト

CRAN (The Comprehensive R Archive Network)

<https://cran.r-project.org/index.html> (日本国内にミラーサイトあり)

関連授業科目

「情報の世界」「統計概論」

その他

なし

担当教員の実務経験

企業の研究所において、データに基づく投資技術の開発に携わりました。本講義では、データサイエンスの基本的な枠組みについて紹介します。データ分析に関する方法を紹介することとあわせて、企業のデータ分析の現場ではどのようなことが行われているかという実際面についても紹介できれば、と考えています。