

データサイエンスの情報技術

必修 2単位

小林 靖之

1. 授業の概要(ねらい)

総合データ応用プログラム修士課程のディプロマポリシーの1つである「データサイエンスに関する基礎的な知識・技能をデータの分析に活用できる」を達成するため、データサイエンスの基礎的な知識・技能を修得する科目の1つです。オンライン教材(LMS)とコンピュータ演習を活用して、学生自身の手でデータサイエンスの情報技術に関する各種基本的スキルを体験しながら修得します。データサイエンスの情報技術の初心者でも容易に準備し取り組めるように、Microsoft Excelによる演習を各単元で行ないます。これらにより、学生自身がデータサイエンスに関して継続的に分析する力を身に付けるきっかけを与え、自身の研究やビジネスに活用できるようになります。

2. 授業の到達目標

以下3点の達成を目指します。

- (1) データサイエンスの分析で最低限必要な過程の情報収集や問題解決のスキルを理解できる。具体的には、データサイエンスに必須である数理統計学や多変量解析の基本的概念や手法を理解できる。
- (2) データサイエンスの問題解決に必要なデータを解析し仮説・モデル化や検証評価する過程を体験を通じて身に付ける。具体的には、数理統計学や多変量解析の基本的手法をコンピュータ上で実行し分析できる。
- (3) データサイエンスの解析手法と必要なソフトウェアや情報処理能力を理解し処理能力を身に付ける。具体的には、公的機関等からデータを入力し、自らが立てた仮説をデータサイエンスの手法で検証できる。

3. 成績評価の方法および基準

コンピュータによる総合演習の成果発表40%・提出レポート40%・関係する単元をまとめた分野の複数の宿題20%で評価します。

ただし、受講や演習への取組み態度が不熱心である場合には成績評価の対象としません。

4. 教科書・参考文献

教科書

LMS上にテキストや演習教材を掲示します。

参考文献

馬場敬之 数理統計学の入門者向け

「確率統計キャンパスゼミ 改訂7」 ISBN:978-4-86615-208-0 マセマ出版社

東京大学教養学部統計学教室 編 数理統計学の中級者向け

「基礎統計学 I 統計学入門」 ISBN:978-4-13-042065-5 東京大学出版会

野口博司 多変量解析の入門者向け

「図解と数値例で学ぶ多変量解析入門」 ISBN 9784542601123 日本規格協会

永田靖, 棟近雅彦 多変量解析の中級者向け

「多変量解析法入門」 ISBN 978-4-7819-0980-6 サイエンス社

Trevor Hastie, Robert Tibshirani, Jerome Friedman 多変量解析の上級者向け

“The Elements of Statistical Learning, 2nd. Ed.” ISBN 978-0-387-84858-7 Springer,

著者が本書を紹介しているURL: <https://hastie.su.domains/ElemStatLearn/>

杉山 将 監訳 多変量解析の上級者向け

「統計的学習の基礎」 ISBN 9784320123625

※上記 “The Elements of Statistical Learning, 2nd. Ed.” の和訳版です。 共立出版

総務省統計局 公的機関のオンライン上の統計データベースの例

e-Stat(日本政府統計の総合窓口) <http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do>

(独立行政法人)統計センター 公的機関のオンライン上の統計データベースの例

教育用標準データセット <https://www.nstac.go.jp/SSDSE/>

University of California Irvine (UCI) 公的機関のオンライン上の統計データベースの例

Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

5. 準備学修の内容

各回授業では、予習として直前回に当たるテキストを読んだ上での例題解答(約1時間)、復習として当日回に当たるテキストの復習・テキストの問題または宿題の該当部分解答(約2時間)を要します。

6. その他履修上の注意事項

以下を前提とします。

(1) 大学学部レベルの数理統計学(確率を含む)・線形代数・微分積分学の基本的概念を理解していること。

必要があれば、本プログラムの科目「データサイエンスの数理」で学修したり、学部の相当科目の教科書等を参照してください。

(2) 表計算ソフトウェアであるMicrosoft Excelの基本的操作(セル操作、関数の利用、グラフ作成等)をPC上で実行できること。必要があればWeb上の統計データベースに自分自身でアクセスできること。

上記ソフトウェア以外のサポートを行ないません。表計算ソフトウェアのMicrosoft Excelを準備する際、本学からライセンス貸与するOffice 365を通信教育課程の学生も利用可能ですので、必要ならば事前に教務グループへ相談してください。

(3) データサイエンスの分析を自分自身で行なう必要のある、研究やビジネス等の何らかの課題をもっていること。

(4) e-learningシステムであるLMSを利用できるPC環境を準備すること。

(5) プレゼンテーションに必要なMicrosoft PowerPointなどによるスライド作成と説明の録音もしくはオンライン会議システムを利用できるPC環境を準備すること。

7. 授業内容

【第1回】 記述統計(1):1変数データについてExcelで可視化し、分散などの基礎的統計量をExcelで計算できる。

【第2回】 記述統計(2):複数変数データについてExcelで可視化し、相関係数、偏相関係数などの統計量をExcelで計算できる。

【第3回】 推測統計(1):様々な確率分布、特に正規分布に関してExcelで計算できる。

- 【第4回】 推測統計(2):推測統計での推定と検定に関してExcelで計算できる。
- 【第5回】 単回帰分析(1):最小二乗法など 単回帰分析の原理を理解し、Excelで計算できる。
- 【第6回】 単回帰分析(2):単回帰分析の解析結果の妥当性をExcelで判断できる。
- 【第7回】 多変量解析における線形代数:多変量解析に必要な、行列の演算、逆行列、固有値分解をExcelで計算できる。
- 【第8回】 重回帰分析(1):線形代数を用い表現された重回帰分析の原理を理解し、Excelで解析した結果の妥当性を判断できる。
- 【第9回】 統計モデルの選択手法:統計モデルや変数の選択に有用な情報量基準(例 AIC)や交差検証法をExcelで計算できる。
- 【第10回】 重回帰分析(2):変数選択法、線形式からの拡張モデルや質的変数の重回帰分析における取扱法をExcelで実行できる。
- 【第11回】 高度な重回帰分析(1):2値目的変数を対象とするロジスティック回帰分析をExcelで実行できる。
- 【第12回】 高度な重回帰分析(2):多重共線性の問題を理解し、その解決法の代表例であるリッジ回帰,PLS回帰,LASSOをExcelで実行できる。
- 【第13回】 他の多変量解析手法:クラスター分析や主成分分析をExcelで実行できる。
- 【第14回】 コンピュータ演習(1):受講生自身が興味をもつ課題に関して、Web上の統計データベースを検索し、Excel等で分析し考察できる。
- 【第15回】 コンピュータ演習(2)とプレゼンテーション:受講生自身が興味をもつ課題の分析結果をプレゼンテーションできる。