

データサイエンス基礎

Data Science Fundamentals

薬：K1-41221MY、生命：K1-41221MS

基礎科目 2年／後期 1.5単位 選択科目

科目責任者 富永 大介(薬学教育研究センター／生命情報科学)

■教育目的

データサイエンスのさまざまな手法と実際のPC操作を学ぶ。基礎的な統計解析法と機械学習法について、基本的な解析の原則と手順を追うことで数値データ解析の実践的技能を身に付ける。またゲノムとたんぱく質の一次および高次構造を視覚的に把握し探索する基礎的手法、およびオープンデータによる創薬分子探索の基礎理論と実践的手法を学ぶ。

■学習到達目標

1. データサイエンスの知識を使って解析を行うための実際のPC操作を身に付ける。(知識・技能)
2. 薬学や生命科学に利用できるオープンデータにはどのようなものがあるかを説明できる。(知識)
3. ゲノム配列の構造とたんぱく質の立体構造を自在に表示し、視覚的に把握できるようになる。(知識・技能)
4. SMILES 表記や薬剤分子探索の原理などケモインフォマティクスの基礎を理解する。(知識)
5. たんぱく質と低分子の相互作用の様子を視覚的に把握できるようになる。(知識・技能)

■準備学習（予習・復習）

予習：「基礎統計学」で学んだ概念などを復習しておく。分子間相互作用、ゲノム配列の構造、たんぱく質の立体構造について用語の意味や概要を調べておく。(30分)

復習：講義で行った課題を条件を変えて再試行してみる。(30分)

■授業形態

双方向型授業（ICT活用）、講義

■授業内容

毎回ノートPCを各自で持参し、実際のPC操作を行いながら演習を中心とした講義を行う。

No.	項目	授業内容	備考・SBOコード
1	データ解析の基礎 1	アルゴリズムとプログラミング	
2	データ解析の基礎 2	要約統計量とデータの可視化	
3	データ解析の基礎 3	t検定	
4	データ解析の基礎 4	分散分析	
5	データ解析の基礎 5	線形回帰	
6	データ解析の基礎 6	モデル選択	
7	データ解析の基礎 7	クラスタリングと主成分分析	
8	機械学習の基礎	SVM	
9	生命科学への応用 1	ゲノム情報解析	
10	生命科学への応用 2	たんぱく質情報解析	
11	薬学への応用 1	化合物データベースと機械学習適用の基礎	
12	薬学への応用 2	ドッキングシミュレーションの原理	
13	薬学への応用 3	オープンデータを使った創薬の原理	

■授業分担者

富永大介 (1～8)、杉原稔 (9～10)、広川貴次 (非常勤講師, 11～13)

■課題（レポート、試験等）のフィードバック及び成績評価方法

毎回の講義での課題の評価の合計を基に総合的に評価する。

■教科書

講義支援システム(MY-CAST)に用意する PDF 形式の授業資料を用いる。

■参考書

Rをはじめよう生命科学のための RStudio 入門, A. P. Beckerman, D. Z. Childs, O. L. Petchey 原著, 羊土社 (2019).
東大式 生命データサイエンス即戦力講座, DSTEP 教材作成委員会, 羊土社 (2021).
超入門! スラスラわかる リアルワールドデータで臨床研究, 康永秀生, 金芳堂 (2025).