

# 情報処理応用演習

講師 ソン ジェヒョン 宋 財 法 (同志社大学文化情報学部 助教)  
E-mail [jasong \[\\*at\\*\] mail.doshisha.ac.jp](mailto:jasong [*at*] mail.doshisha.ac.jp)  
Homepage <https://www.jaysong.net>

日時

- 2020 年 8 月 21 日 (金) 2~4 コース
- 2020 年 8 月 24 日 (月) 2~4 コース
- 2020 年 8 月 25 日 (火) 2~3 コース

教室 1 号館・情報処理室

## 1 授業の内容

本講義は「R の使い方」を中心とする。したがって、統計学、確率、多変量解析、因果推論などについては解説しない。データ分析のツールは数多く存在するが、2010 年以降のデータサイエンスの主なツールは R と Python である。しかし、これまでのデータ分析ソフトウェアがマウスクリックや簡単なコマンドで分析が可能であった一方、R と Python はプログラミング言語でもあるため、敷居が高いと考えられる。また、R の場合、`tidyverse` という思想の下で独自の書き方へ発展を遂げている。しかし、R を用いたデータ分析の教科書の多くはある程度 R の前提知識を持つ読者層を想定するが多い。また、R の基本的な使い方が含まれる教科書の場合も、付録レベルでしか触れていないのが現状である。この場合、教科書の R コードを読んでも「なぜこのコードか？」の疑問が生じやすい。本講義の目標は R を用いたデータサイエンスの教科書が読める土台づくりとも言えよう。

## 2 授業の目標

本講義の到達点は以下の 3 点である。

- R Markdown の使い方を習得する
- 入手したデータを分析に適した形へ整形する方法を習得する
- データ分析に頻繁に用いられるグラフを作成する方法を習得する

### 3 評価

- 平常点 20%: 授業への参加度、質問など (講義外における活動を含む)
- 期末課題 80%: 期限: 2020 年 8 月 30 日 (予定)

### 4 履修上の注意

講義開始日から期末課題の締切日まで随時質問を受け付ける。

- 講義日
  - 講義中および前後で質問を受け付ける
- 講義外
  - メールまたは Zoom による対応
  - 講師が 24 時間対応できないため、Zoom による対応を希望する場合、事前に宋までメールすること。Zoom のルーム ID はメールで知らせる。

### 5 教科書・参考書

教科書は用いず、スライドのみで対応する。以下の参考資料は本講義内容を全て、または一部カバーする。

- 『私たちの R: ぼくらがかんがえたさいきょうの R ほん』
  - <https://www.jaysong.net/RBook/>
  - SONG と矢内勇生 (高知工科大学) が共同執筆中の E-Book
  - 本講義の内容を 100% カバーする
- 浅野正彦・中村公亮. 2018. 『はじめての RStudio: エラーメッセージなんかこわくない』オーム社
  - 初心者向け
- 松村優哉 他. 2018. 『R ユーザのための RStudio[実践] 入門—tidyverse によるモダンな分析フローの世界—』技術評論社
  - 中級者向け

## 6 講義内容・参考文献

### 6.1 1日目: R と RStudio の基本的な使い方

- R と RStudio の導入
- R Markdown の基礎
- 電卓としての R
- 表形式データの読み込み/書き出し

### 6.2 2日目: 表形式データの操作

- 行と列の抽出
- データの拡張
- データの要約
- データの結合
- 整然データの作成

### 6.3 3日目: 可視化

- ggplot2 の仕組み
- 棒グラフ
- ヒストグラム
- 箱ひげ図
- 散布図
- 折れ線グラフ

最終修正: 2020 年 8 月 4 日