

第1章: イントロダクション

Elena Llaudet & Kosuke Imai.

Data Analysis for Social Science: A Friendly and Practical
Introduction.

2026-03-09

1.1 本書の概観

本書の目的と構成

- ▶ 社会科学におけるデータ分析の基本を、予備知識なしの状態から学ぶ。
- ▶ 統計ソフト R を用いて現実世界のデータを分析する。
- ▶ 計量社会科学研究の 3 つの基本的な目標を学ぶ：
 1. 測定 (Measurement)
 2. 予測 (Prediction)
 3. 説明 (因果推論、Causality)

各章の概要

- ▶ 第 1 章：イントロダクション（R の基本操作）
- ▶ 第 2 章：無作為化実験による因果効果の推定（説明）
- ▶ 第 3 章：社会調査研究による母集団の特徴の推論（測定）
- ▶ 第 4 章：線形回帰を用いた結果の予測（予測）
- ▶ 第 5 章：観察データによる因果効果の推定（説明）
- ▶ 第 6 章：確率
- ▶ 第 7 章：不確実性の数値化

1.4 なぜデータ分析を学ぶのか?

データ分析スキルの重要性

- ▶ 社会学者として、母集団の特徴を測り、予測を行い、因果関係を解明するためにデータが必要。
- ▶ 意思決定者に情報を提供するデータアナリストの需要が高まっている。
- ▶ ビッグデータ時代において、計量研究に精通することは非常に価値のあるスキル。

コードの学習

- ▶ コンピュータに特定の命令（プログラミング言語）を与えることで分析を実行する。
- ▶ 本書では、世界中で広く利用されている統計プログラミング言語 **R** を使用する。

1.5 準備 (R と RStudio)

R と RStudio

- ▶ **R**: 統計計算とグラフィックスのためのソフトウェア (エンジン)。
- ▶ **RStudio**: R をより使いやすくするためのインターフェース (IDE)。

作業ディレクトリの設定

- ▶ データセットを読み込む前に、R にファイルの場所を教える必要がある。

```
# 作業ディレクトリの設定例
```

```
# setwd("~/Desktop/DSS")
```

1.6 R の紹介

R で計算をする

- ▶ R を計算機として使用できる。

```
1 + 3
```

```
## [1] 4
```

オブジェクトの作成

- ▶ 情報を「オブジェクト」として保存する (<- 割り当て演算子)。

```
four <- 1 + 3
```

```
four
```

```
## [1] 4
```

```
hello <- "hi, nice to meet you"
```

```
hello
```

```
## [1] "hi, nice to meet you"
```

関数の使用

- ▶ 関数は特定の操作を実行する (関数名 (引数))。

```
sqrt(16)
```

```
## [1] 4
```

```
sqrt(four)
```

```
## [1] 2
```

1.7 データを読み込んで理解する

データの読み込み (1)

- ▶ `read.csv()` 関数を用いて、外部のデータセットを R に読み込みます。

1. ローカルに保存した *STAR* データの読み込み (推奨)

```
star <- read.csv("STAR.csv")
```

(参考) URL から直接読み込むことも可能

```
# star <- read.csv("https://ayumu-tanaka.github.io/QSS/DSS_Data/STAR.csv")
```

データの理解 (2)

▶ 読み込んだデータの中身や構造を確認します。

1. 最初の 6 行を表示

```
head(star)
```

```
##   classtype reading  math graduated
## 1    small     578   610         1
## 2  regular     612   612         1
## 3  regular     583   606         1
## 4    small     661   648         1
## 5    small     614   636         1
## 6  regular     610   603         0
```

2. 次元の確認 (行数, 列数)

```
dim(star)
```

```
## [1] 1274    4
```

変数の種類

- ▶ **文字変数** (Character variables): テキストデータ。
- ▶ **数値変数** (Numeric variables): 数字。
 - ▶ **二値変数** (Binary variables): 0 か 1 の 2 つの値のみ (ダミー変数)。
 - ▶ **非二値変数** (Non-binary variables): それ以外の数値 (例: テストの点数)。

1.8 平均値を計算し、解釈する

変数へのアクセス

- ▶ \$ 記号を用いて、データフレーム内の特定の变数にアクセスする。

```
# reading 変数の最初の数値を表示  
head(star$reading)
```

```
## [1] 578 612 583 661 614 610
```

平均値の計算

▶ `mean()` 関数を用いて平均値を計算する。

```
# 読解力テストの平均点
```

```
mean(star$reading)
```

```
## [1] 628.803
```

```
# 卒業率（二値変数）の平均
```

```
mean(star$graduated)
```

```
## [1] 0.8697017
```

解釈

- ▶ 二値変数の平均値に 100 を掛けると、その特徴を持つ割合 (%) になる。
- ▶ 例: `graduated` の平均が 0.87 であれば、約 87% の生徒が卒業したことを意味する。

1.9 まとめ

第1章のまとめ

- ▶ R と RStudio の基本的な操作を学んだ。
- ▶ データの読み込み方法、データフレームの構造、変数の種類を理解した。
- ▶ `mean()` 関数を用いて記述統計量（平均値）を算出し、その解釈方法を学んだ。
- ▶ 次章からは、いよいよ実際の研究例を用いて因果関係の推定（説明）に取り組む。