石井一夫（公立諏訪東京理科大学）

データサイエンス入門 教養としてのデータサイエンス

北川源四郎，竹村彰通 編
内田誠一，川崎能典，孝忠大輔，佐久間淳，椎名洋，中川裕志，樋口知之，丸山　宏 著
講談社サイエンティフィク（2021），1,980 円（税込），240p., ISBN：978-4-06-523809-7

モデルカリキュラムリテラシーレベルのテキストとして

本書籍は，「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム」（以下，モデルカリキュラム）に準拠し，認定が進められている「数理・データサイエンス・AI 教育プログラム認定制度（リテラシーレベル）」のテキストとして使用することを目的としたものであると考えられる．本書籍の構成は，モデルカリキュラムと完全に一致しており，第 1 章［導入］社会におけるデータ・AI 利活用，第 2 章［基礎］データリテラシー，第 3 章［心得］データ・AI 利活用における留意事項の 3 章構成となっている．

モデルカリキュラムにおいてリテラシーレベルは，文系理系共通の導入科目の位置づけであると思われる．本書籍においても，統計学を学ぶにしても，プログラミングを学ぶにしても，何かを学ぶというよりは，ガイダンス的な内容で広く浅く説明している．したがって，本書籍を読めばデータサイエンスとは何か，どういう内容か，それで何ができるのかをさわりだけ触れることはできるが，データ分析ができるようになることは難しい．一方で，書籍そのものは易しいので，さらっと読んで概要をつかみ，興味が出てきた項目についてはほかの資料を読み進む本と考えるとよい．

社会におけるデータ・AI 利活用

「1.1 社会で起きている変化」，「1.2 社会で活用されているデータ」，「1.3 データと AI の活用領域」，「1.4 データ・AI 利活用のための技術」，「1.5 データ・AI 活用の現場」，「1.6 データ・AI 利活用の最新動向」と，モデルカリキュラムの項目に沿って，キーワードの解説がされている．ビッグデータや AI の利活用促進に基づく第 4 次産業革命や Society 5.0 による社会の変化が紹介されており，そこで利用されているデータや応用分野，応用技術についての最新動向の，用語も含めた解説が試みられている．モデルカリキュラムがまとめられたのがコロナ禍の前のことであり，最近ますます問題が大きくなっている気候変動・地球温暖化をはじめとする SDGs（Sustainable Development Goals）については，本書籍では触れられておらず，テレワークを中心としたニューノーマルな生活様式や，脱炭素社会に向けたデータサイエンスの在り方についても触れられていない．しかし，これらは，従来のものづくりを中心とした産業や生計の在り方や，シンギュラリティなど AI と人間との関係にも再考を強いるものであり，今後考慮する必要があると思われる．

データリテラシー

第 2 章は，「2.1 データを読む」，「2.2 データを説明する」，「2.3 データを扱う」という，実際にデータと接する局面での対応について，データとのかかわりを基に書きつづられている．まず，「2.1 データを読む」では，データの種類，データの分布と代表値，ばらつきと誤差の扱い，相関と因果関係，母集団と標本抽出，データセットの扱いや統計の考え方など統計の基礎に関する項目について説明がなされている．もっとも，数式的な扱いなどは触れられておらず，雲をつ

かむような感じがするのは否めない．統計学についてよく理解したい方は，ここをベースにより専門の書籍を手にすることをお勧めする．「2.2 データを説明する」では，グラフ（ヒストグラム，棒グラフ，円グラフ，散布図，ヒートマップなど）を用いた表現とその解釈，取り扱いについて述べられている．「2.3 データを扱う」では，表形式のデータ，すなわちスプレッドシートのデータの説明とその扱い，たとえば，データの集計や，並べ替え，ランキング，散布図の描き方などの説明である．いわゆるアプリケーションソフトを使ってデータを眺めたり，スプレッドシートを使った簡単な処理について，要点を説明するにとどめている．分析者の使う統計ツールには R や Python があるが，本章では R が要点だけ紹介されている．本章だけでは，R や Python を使いこなすというレベルには到底到達し得ないので，これらのツールを使いこなしたい方は，より専門的な書籍に当たる必要がある．

データ・AI 利活用における留意事項

第 3 章は，「3.1 データ・AI を扱う上での留意事項」，「3.2 データを守る上での留意事項」からなる．「3.1 データ・AI を扱う上での留意事項」は，まず，ELSI（Ethical, Legal and Social Issues），すなわち，倫理，法律，社会的側面から，データを扱う上での留意事項について述べている．次に，データの品質保証の問題を取り上げている．具体的には，AI のブラックボックス化，説明可能性，アカウンタビリティ，透明性，公平性などの問題である．データの品質保証の問題については，データの正確性を担保する上でも重要な事項であり，注意を払う必要がある．これらは，最近の AI の運用上でも話題になっている事項である．「3.2 データを守る上での留意事項」は，データサイエンス上の情報セキュリティとプライバシーの問題を取り上げている．情報セキュリティについては，機密性，完全性，可用性の 3 点から述べられており，プライバシーについては，プライバシーの定義と匿名化，匿名加工の問題について述べられている．特に，ビッグデータを扱う場合には，患者データや顧客データなど個人情報に配慮しなければならない場面は増えており，今後ますます重要性が増してくると思われる．

どういう人に薦めるか

AI，ビッグデータ，データサイエンスで何が話題になっているかを知って，これから何を学ぶか決めたい人や，データサイエンスの全容をざっくりとつかみたいという人，特に高校生や，大学生の初学年の人には向いている．また，「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム」や，「数理・データサイエンス・AI 教育プログラム認定制度（リテラシーレベル）」の活用や運営にあたって，どのようなことが教えられているかを知りたい大学教員や，企業の人事担当者，データ分析部門の担当者などにお薦めする．しかし，これだけで，データサイエンスを学んだというには，あまりに薄い内容であるので，より高度な知識を身に付けたい学生には，ここをベースにほかの書籍を当たって知識を深めてほしい．

追記）本記事の執筆後，データサイエンスに大きなトレンドが見られるので，強調しておきたい．本文中でも少し触れているが，SDGs で表されるような社会的重要課題の解決への手段と指針としてのデータサイエンスの重要性が高まっていることである．具体的には，コロナ禍や，気候変動，少子高齢化などの地球規模や国家規模の課題へどう対峙していくかという側面である．これは，人類の叡智と行動力が試される厳しい選択を迫られるものになりそうである．

（2021 年 8 月 9 日受付）

石井一夫（正会員）
kishii@rs.sus.ac.jp
公立諏訪東京理科大学工学部情報応用工学科教授，久留米大学医学部内科学講座心臓・血管内科部門客員准教授．専門分野：ビッグデータ分析，計算機統計学，データマイニング，数理モデリング，機械学習，人工知能．医療ビッグデータ，気象ビッグデータ研究に従事．2015 年度本会優秀教育賞受賞．日本技術士会フェロー，APEC エンジニア，IPEA 国際エンジニア．