

教養としてのデータ・サイエンス教育 Moocs の活用を視野に入れて 日本数学会教育委員会シンポジウム@金沢大学

鈴木寛 (Hiroshi Suzuki)

国際基督教大学 (International Christian University)

2019年9月17日

講演内容項目 (Contents)

- ① はじめに (背景・個人的な経験から)
- ② 研究者のためのデータ活用と分析 (Data Analysis for Researchers)
- ③ 教養としてのデータ・サイエンス
- ④ Massive Open Online Courses (MOOCs) から学ぶ
- ⑤ 「教養としてのデータ・サイエンス」のコースを企画するにあたって
- ⑥ おわりに

背景・現状 I

数理・データサイエンス

政府・文部科学省

- 第5期科学技術基本計画：2016年1月22日閣議決定^a
<https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>
- 大学の数理・データサイエンス教育強化方策：2016年12月21日公表
 - 文部科学省として喫緊に取り組むべき方策
 - ① 数理・データサイエンス教育研究センター (仮称) の整備
 - ② 標準カリキュラム・教材の在り方
 - ③ 実践教育に関する産学連携ネットワークの整備
- <https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>
 - 「数理及びデータサイエンスに係る教育強化」の拠点校の選定
北海道・東京・滋賀・京都・大阪・九州 (6 国立大学)
- データ関連人材育成プログラム (D-DRIVE^b) 2017年度～
- 「大学における数理・データサイエンス教育の全国展開」の協力校の選定
 - 20 大学選定：2019年1月8日
 - http://www.mext.go.jp/b_menu/shingi/chousa/koutou/095/gaiyou/1412367.htm

^a第2章 未来の産業創造と社会変革に向けた新たな価値創出の取組

^bDoctoral program for Data-Related Innovation Expert

背景・現状 II

数理・データサイエンス・AI

政府・文部科学省

- 経済産業省：産業技術環境局 大学連携推進室等 2019年3月26日
 - 数理資本主義の時代～数学パワーが世界を変える～
https://www.meti.go.jp/shingikai/economy/risukei_jinzai/20190326_report.html
- 統合イノベーション戦略推進会議 (首相官邸・政策会議)
 - 人間中心の AI 社会原則：2019年3月29日
<https://www8.cao.go.jp/cstp/aigensoku.pdf>
 - AI 戦略 2019～人・産業・地域・政府全てに AI～ 2019年6月11日
<https://www.kantei.go.jp/jp/singi/tougou-innovation/pdf/aissenryaku2019.pdf>
- ④ 目標 1：文理を問わず、全ての大学・高専生 (約 50 万人卒/年) が、課程にて初級レベルの数理・データサイエンス・AI を習得 [MOOC 放送大学の活用]
- ⑤ 目標 2：多くの社会人 (約 100 万人/年) が、基本的情報知識と、データサイエンス・AI 等の実践的活用スキルを習得できる機会をあらゆる手段を用いて提供
- ⑥ 目標 3：大学生、社会人に対するリベラルアーツ教育の充実 (一面的なデータ解析の結果や AI を鵜呑みにしないための批判的思考力の養成も含む)

背景・現状 III 標準カリキュラム・教材 (2019 秋)

数理・データサイエンス教育強化拠点コンソーシアム

- 北海道大学: <https://www.mdsc.hokudai.ac.jp>
- 東京大学: <http://www.mi.u-tokyo.ac.jp>
エクステンション株式会社 <https://www.utokyo-ext.co.jp/>
- 滋賀大学: <https://www.ds.shiga-u.ac.jp/dscenter/>
- 京都大学: <http://ds.k.kyoto-u.ac.jp>
- 大阪大学: <http://www-mmds.sigmath.es.osaka-u.ac.jp>
- 九州大学: <http://mdsc.kyushu-u.ac.jp>
- 神戸大学: <http://www.cmds.kobe-u.ac.jp>
- 北見工業大学・東北大学・山形大学・筑波大学・宇都宮大学・群馬大学・千葉大学・お茶の水女子大学・新潟大学・長岡技術科学大学・静岡大学・名古屋大学・豊橋技術科学大学・島根大学・岡山大学・広島大学・愛媛大学・宮崎大学・琉球大学

教科書の出版など

- データサイエンス入門シリーズ、講談社、2019 年秋刊行開始

データ・サイエンス

教育と教育 IR (Institutional Research) と個人的な学び

教育: 研究者のためのデータ活用と分析 (Data Analysis for Researchers)

- 全学の大学院 (修士) の学生向け、学部 3 年生以上は許可を得て履修可
- 受講生: Rotary Peace Fellow, The Project for Human Resource Development Scholarship (JDS) および一般の大学院生と少数の 4 年生 10-25 人
- 教授言語: 英語
- スケジュール: 2 時限 (70 分 × 2 × 10 週間)、1 時限講義・1 時限実習
- 担当教員: 経済学、情報科学、数学 (R Markdown etc.)
- 担当年度: 2014-2015^a, (2016), 2017 年

^a2015 年度までは、研究者のためのコンピューティング (Computing for Researchers)

教育 IR

学修・教育センターの柱の一つに、大学の IR とは別に、教育 IR を含める

個人的な学び

2019 年 3 月末に退職後 MOOCs のコースなどを利用して、学習中

「研究者のためのデータ活用と分析」について

Data Analysis for Researchers

スケジュール

- 1 Introduction to R, Open Data and Free Software
- 2 Basic R Objects and Commands
- 3 Data Frame Manipulation
- 4 Linear Regression and Graphics
- 5 Dynamic Documents Using Rmarkdown
- 6 Statistical analysis with R II
- 7 Statistical analysis with R III
- 8 Statistical analysis with R IV
- 9 Guest Lecture and preparation for presentations
- 10 Final presentations

コースの特徴

オープン・データ、R Studio (サーバーと PC) 上での R の活用、R Markdown での文書作成、実例の提供、ゲスト講義、プロジェクト、教員の協力

「研究者のためのデータ活用と分析」について

Data Analysis for Researchers

授業で使ったデータ

- Base R に付属のデータ
 - cars: 車のスピードと制動距離
 - iris: あやめの花びらと萼のデータ
- package MASS に付属のデータ
- WDI: World Bank Development indicators for R


```
library(WDI)
#GDP (current US$)
gdp <- WDI(country = c("US", "JP", "CN", "KR"),
indicator = "NY.GDP.MKTP.CD",
start = 1960, end = 2017)
```

 現在は、より新しい wbstats パッケージもある。
- Quandl package: <https://www.quandl.com/tools/r>
- Google Trends: <https://trends.google.co.jp/>
- Yahoo Finance: <https://finance.yahoo.com/quote/DATA/>

教養としてのデータ・サイエンス

一般の学生対象のコース

なぜ 社会的要請? 教養人としてのリテラシー

- 科学的思考リテラシー：因果関係、相関関係、隠れた因子についての考察
- 関係する因子の多い実社会の問題の分析・解決方法の探索
- 社会や会社・機関の責任ある構成員として、政策決定や事業方針策定に必要な因子を把握・異なる視点から検証
- 関係するすべての人が知恵を出し合い、問いを発し、異なる視点を理解し、協力し、新しい価値を生み出しながら、データ・サイエンス、AI を活用
- データの公開が進み、だれでも、膨大なデータを活用することが可能
- 変化の時代に、データの分析結果を理解し、適切な判断をすることが不可欠
- 数学への苦手意識が強くても、データ分析に力を発揮するひとは多い
- (高校までに学ぶ) 数学を活用し、その重要性を実感することが可能

コース・デザイン

- 目的を達成できるコースの企画・運営が不可欠
- 数学の専門知識の活用を急がず、(変化の時代に) 数学を学ぶ有効性を探る

日本での「教養としてのデータ・サイエンス」

課題 (私見) 皆さんも様々なお考えがあると思います

- データの公開とその利用が進んでいない
- 高校までの教育でデータ・サイエンスが扱われてこなかった
- 特に学部では日本語の閉じた空間で学んでいる場合が多い
- 学部レベルでの学際的な科学教育が進んでいない
- 数学分野の問題 (他の分野でもそうかもしれませんが)
 - 授業で、コンピュータを活用することが少ない
 - 授業で、応用や社会とのつながりに触れる機会が少ない
 - 統計学の授業が実際のデータを扱うデータ・サイエンスに連動していない
 - 数学分野の教員のデータ・サイエンス (教育) の理解度が低い? どのような数学がデータ・サイエンスに必要なかという意識しかない? 数学の教員としてどのように貢献したら良いかわからない。
- 理系の学生は (文系の学生も)、就職先でデータ・サイエンスや AI を活用することが多いが、大学のカリキュラムは、学生の必要に応えていない
- 学部を横断して、文系 (特に社会科学系) の教員と協力して、教える機会がない
- 自分の分野の研究、授業、大学のしごとで、手一杯で、それ以上のことは考える余裕がない
- **大学全体で取り組み、数学教員が少しずつ役割を見出していく価値がある**

鍵となるもの I

Teaching to Learning

教養としての DS

動機づけ・自律的学習・協働と評価

理想的学修者モデルを想定 教養? Liberal Arts?

- 強い動機づけを持ち、自律的かつ主体的に、一生学び続ける基盤を構築
- 個別学習では得られないことをグループの中で積極的に学ぶ
- 広い興味をもち、自らの地平を広げながら学ぶ
- 複雑な問題の全体像を把握し・課題設定し・他者と協力しながら価値を創造

クラス内での学びの機会の提供 Teaching to Learning

- 動機づけ、世界が広がっていくための機会と情報の提供
- 何をグループや、クラスで行うか。
 - 自律的に学び始めるため、スタート地点に立つ機会を提供
 - 問い (課題) を提出する訓練 Community of Inquiry (Col),
 - 可視化の評価 Communication of Facts

学びの支援 Students with Various Backgrounds

- サポートの学生の活用と訓練：質問を受けたことのレポート
- 教員の役割・(コースの目的に適した) 評価

鍵となるもの II

Inclusive Education・Universal Access

教養としての DS

世界の様々な課題に向き合うために

Literacy Expand your horizon!

- 英語：学びの広がりを制限しないため
 - データの取得、コマンド、プログラミング
 - マニュアル、ヘルプ、Q & A、(ユーザ) コミュニティ へのアクセス
 - オンライン・コース、オンライン・ツールの利用
- 数学：学びの広がりを制限しないため
 - 基本的な式変形など、中学数学および、高校数学 I, II, A, B の必要な部分の学習
 - 確率・統計：データ・サイエンスの中で少しずつ学習
 - 理系で数学を学んだ学生が、教え・助ける経験をもつ
- コンピュータ：学びに有効活用するため
 - データ・サイエンスのためのソフトによるスクリプトの活用
 - プログラミングへの関心

Resources IT / Cloud

- オープン・データ (Open/Public Data)
- インターネット上の情報、コース、各種ツールなど (Online/Cloud)
- 無償、かつ、オンラインのソフトウェア (Free and online/cloud system)

総務省：オープン・データの定義

- ① 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
- ② 機械判読に適したもの
- ③ 無償で利用できるもの

http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/

World Bank: Open Data Defined <http://opendatatoolkit.worldbank.org>

The term "Open Data" has a very precise meaning. Data or content is open if anyone is free to use, re-use or redistribute it, subject at most to measures that preserve provenance and openness.

- ① The data must be legally open, which means they must be placed in the public domain or under liberal terms of use with minimal restrictions.
- ② The data must be technically open, which means they must be published in electronic formats that are machine readable and non-proprietary, so that anyone can access and use the data using common, freely available software tools. Data must also be publicly available and accessible on a public server, without password or firewall restrictions. To make Open Data easier to find, most organizations create and manage Open Data catalogs.

オープン・データの例

List of Open Data Catalogue

- 公共データ：<https://www.data.go.jp>
オープンデータの意義・目的：1. 国民参加・官民協働の推進を通じた諸課題の解決、2. 経済活性化行政の高度化・効率化、3. 透明性・信頼の向上
 - データベースサイト一覧：<https://www.data.go.jp/list-of-database/>
 - 気象庁：<https://www.jma.go.jp/jma/menu/menureport.html>
- U.S. Government's Open Data: <https://www.data.gov>
- EU Open Data Portal: <http://data.europa.eu/euodp/en/home>
- UK Open Data: <https://data.gov.uk>
- World Bank: *New Ways of Looking at Poverty*
 - Open Data: <https://data.worldbank.org>
 - World Development Indicators: <http://datatopics.worldbank.org/world-development-indicators/>
- UN Data: <http://data.un.org>
- WHO Data: <https://www.who.int/gho/en/>
- Google Public Data: 日本語での検索：7、言語を英語に変えると：136
<https://www.google.com/publicdata/directory>
- Open Knowledge Foundation: <https://okfn.org>
 - Global Open Data Index: <https://index.okfn.org>

Free Software, Online Access

R

- R Project for Statistical Computing: <https://www.r-project.org>
- R Studio: <https://www.rstudio.com>
- R Studio Cloud: <https://rstudio.cloud>

Python

- Python: <https://www.python.org>
- Anaconda: <https://www.anaconda.com>
- Jupyter Notebook Cloud: Binder, Kaggle Kernels, Google Collaborate, CoCalc, PaizaCloud, etc.

Free Software

- Free Software, Free Society: Selected Essays of Richard M. Stallman: <https://www.gnu.org/philosophy/fsfs/rms-essays.pdf>
- Richard Stallman TEDxGeneva 2014: https://youtu.be/Ag1AKI1_2GM

インターネット上での学習

Online Learning Source

List of Online Help and Mini Courses

- 英語では、Online で学べるサイトが多い。基本コースは無料。以下は例
 - TutorialPoint: <https://www.tutorialspoint.com/>
 - DataCamp: <https://www.datacamp.com/home>
 - Code Academy: <https://www.codecademy.com>
 - RStudio Premier: <https://rstudio.cloud/learn/primers>
- 質問とその答えの載ったサイト、User Community

MOOCs

大学などが提供している大規模公開オンライン講座

- OED: MOOC n. massive open online course, an educational course made available to a large number of people via the internet.
- First MOOC: 2008 by Dave Cormier, Connectivism and Connective Knowledge (CCK08)
- MIT OpenCourseWare 2002: 教育の質の向上（教育の提供ではない）OER
Stanford U Model: 有名教授の授業、UC Berkeley: 学習支援
- MOOC 元年: 2012年 Stanford U. と MIT がオンラインのコースを提供

Moocs の活用と注意点

MOOCs の活用を勧める理由

英語の数学の教科書と似た利点

- 無償で、講義ビデオ、講義録、スクリプトがダウンロード可能な場合が多い
- 基本的な内容が適切に学習できるように、配置されている
- いくつかのコースからなるプログラムも多くあり、修了証を提供
- 例・利用する データ (datasets) が優れている (個人で準備するのは困難)
- プログラムを含む演習問題が内容を理解する助けになっている

MOOCs 活用時に考えるべきこと

学修者に適したコースの推薦

- 無償でコンテンツにアクセスできるコースが多いが 修了証のため有償のプログラムを勧められる
- 演習問題によって、学びの確認ができるが、無償では、すべての演習問題ができない、または利用期間が限られている場合が多い
- ビデオ講義の内容も重要だが、演習問題の質が学びには重要、パスできないときの方策も必要
- 同じものが何回も利用されている。洗練もされるが、修了証の質を維持するためか、コースによっては、簡単には答えられないようにした問題もある

Moocs の データ・サイエンス・コースの例

Coursera

Stanford U. が中心

- Stanford U.: Machine Learning 人気の高いコース (登録: 2,513,476)
<https://www.coursera.org/learn/machine-learning>
- Johns Hopkins U.: Data Science, 10 courses [初級, R]
<https://www.coursera.org/specializations/jhu-data-science>
- U. of Michigan: Applied Data Science, 5 courses [中級, python]
<https://www.coursera.org/specializations/data-science-python>
- 100% オンラインのデータ・サイエンスの修士号取得のコースもある。
U. of Illinois, U. of Michigan, U. of Colorado, など

edX

MIT と Harvard U. が中心

- Harvard U.: Data Science, 9 courses [初級, R]
- MIT: Statistics and Data Science, 5 courses [Graduate Level]
<https://www.edx.org/micromasters/mitx-statistics-and-data-science>
- Microsoft, IBM, UC SanDiego などコースをプログラムを提供

Professional Certificate in Data Science

基本情報: HarvardX, through edX

演習も無償で殆どすべて提供

- URL: <https://online-learning.harvard.edu/series/professional-certificate-data-science>
- Book: <https://rafalab.github.io/dsbook/> (R Markdown Document)

R で学び、Data Camp で Assessment

① Data Science: R Basics; データ解析ソフト R の基本	8 weeks
② Data Science: Visualization; データの視覚化	8 weeks
③ Data Science: Probability; 確率・大数の法則	8 weeks
④ Data Science: Inference and Modeling; 推定と数学モデル	8 weeks
⑤ Data Science: Productivity Tools; Unix, Git, GitHub, R Markdown	8 weeks
⑥ Data Science: Wrangling; データの整理	8 weeks
⑦ Data Science: Linear Regression; 線形回帰	8 weeks
⑧ Data Science: Machine Learning; 機械学習	8 weeks
⑨ Data Science: Capstone まとめと次のステップへの架け橋	2 weeks

コースの前半で使われるデータ例

Professional Certificate in Data Science

Required R Packages for Examples: tidyverse, dslabs:

<https://cran.r-project.org/web/packages/dslabs/dslabs.pdf>

- 男女の身長データ
- アメリカにおける 2010 年の銃による殺人事件の FBI の報告書
- Gapminder: Almost nobody knows the basic global facts! (Gapminder Test)
 - TED 増え続ける世界人口 (Hans Rosling)
<https://www.gapminder.org>
 - Health and income outcomes for 184 countries from 1960 to 2016
 - Country, Year, Infant deaths per 1000, Life expectancy in years, Average of children per woman, Country population, GDP, Continent, Geographical region
- Brexit の国民投票の事前調査と実際のデータ
- 2016 年の大統領選挙の事前調査と結果のデータ
- UC Berkeley の大学院入試における男女差
- オランダにおける研究費獲得率に関する男女差
- プロ野球の新人の打率予測

大学生のためのデータサイエンス (1) の内容

gacco: ドコモ提供

- Week 1 現代社会におけるデータサイエンス
 - 役割
 - データの取得・管理、入手方法、分析
 - 画像・音声処理技術
- Week 2 データ分析の基礎
 - ヒストグラム・箱ひげ図・平均・分散・標準偏差・散布図・相関係数・回帰直線
 - データ分析で注意すべき点：相関と因果、観察研究と実験研究、標本調査
- Week 3 コンピュータを用いたデータ分析
 - Excel, R, Python の紹介
- Week 4 データサイエンスの応用事例
 - 保険・金融
 - マーケティング・リサーチ
 - 染色体上で遺伝子を探す
 - 品質管理

JMOOC: <https://www.jmooc.jp>

10月8日開講

統計学 II: 推測統計の方法 (竹村先生他)・社会人のためのデータサイエンス入門

全学の学生のためのコースの企画について

配慮すべきこと

学習目標

- オープン・データについて理解し、課題を設定し、活用することを経験
- データ分析ソフトに触れ、統計量について理解し、分析と可視化を経験する
- データ分析のプロセスの全体像と流れを理解し、そのステップを経験する
- データの分析結果の見え方の評価を受ける
- 何が見えて、なにが見えないか、どのようなものが必要かともに議論し改善していく経験をもつ
- あることを知るためには、どのようなデータが必要か、取得したデータから読み取れるものはなにかを議論する

R Markdown

Or jupyter notebook for python

- Reproducible Research: 再現性を確保でき、記録としても有用
- Literate Programming: Code, Script の説明を順次書き加えることができる
- html, doc, presentation, pdf (using TeX) フォーマットに出力できる
- スクリプトの利用、演習の記録、課題の提出、プレゼンテーションにも有用

全学の学生のためのコースの企画について

配慮すべきこと

考慮すべきこと

問い

- 自律的・主体的な学修者が、大学のコースとしてデータ・サイエンスを学ぶ価値は何か。必修要件を修了証で代替できるか
- 学修者が助け合い、学び合い、智を、創造していくためには何が必要か
- 学修者に、どのような支援が、必要で、有効か
- Empirical (理論ではなく、実験 [実証, 経験] に基づいた) 面が大きいことを意識して、理論より例の提示を重視し、背景にある数学を説明
- 社会科学系など異分野の教員と助け合い、協働でコースを運営

変化の時代に、多くの因子が関係する課題への数学の貢献を考えながら

Quote: On Listening to Lectures, by Plutarch

プルタルコス

The correct analogy for the mind is not a vessel that needs filling, but wood that needs igniting - no more - and then it motivates one towards originality and instills the desire for truth. (<https://quoteinvestigator.com/2013/03/28/mind-fire/>)
 ころは、満たすべき器ではなく、着火すべき木のようなものなのだ。火がつけば、あとは自分で、どんどん、真理を吸収していくことだろう。(私訊)

データ・サイエンス教育と大学の未来

Quote:

Apple co-founders Steve Jobs and Steve Wozniak didn't have degrees when they launched what has become one of the most valuable companies in the world. And now Apple CEO Tim Cook is spreading the word that would-be programmers really don't need the endorsement of a university to be able to create something of commercial value, such as an app for the Apple App Store.

<https://www.zdnet.com/article/apple-ceo-tim-cook-you-dont-need-a-degree-to-code-mobile-apps/>

ひとこと

データ・サイエンスや AI に関して、大学で教えるべきことは何なのでしょう。今日は「教養としてのデータサイエンス教育」というタイトルでお話しましたが、どのレベルであっても、真剣に考えるべき課題だと思います。大学以外で学べるのであるならば、大学の価値は、失われていくかもしれません。このことを考えながら、準備をしました。みなさんが、一般教育科目、数学の学生の教養として、学内全体のデータ・サイエンス教育のことを検討するときに、なにかヒントになることが含まれていればと願っています。