

データサイエンス教育の滋賀大学モデル

竹村 彰通[†]・和泉 志津恵^{†*}・齋藤 邦彦[†]・
姫野 哲人[†]・松井 秀俊[†]・伊達 平和[†]

(受付 2017年6月28日; 改訂 9月25日; 採択 10月4日)

要 旨

ビッグデータから価値を引き出すための方法がデータサイエンスである。データサイエンスの技術的基礎をなすのは、データエンジニアリング(情報学)とデータアナリシス(統計学)である。理系のスキルを持つとともに文系的なマインドも持つデータサイエンティストを育成するため、滋賀大学データサイエンス学部が2017年4月に誕生した。本論文では、滋賀大学のデータサイエンス教育の内容を解説する。まず滋賀大学モデルの概要をまとめる。続いて、データサイエンスの情報系と統計系のカリキュラムについて、それぞれ述べる。そして、これらのカリキュラムをサポートする学習環境について述べる。さらに、課題解決型(Project-Based Learning, PBL)演習のための国内の組織との連携の実例を幾つか紹介する。国内の多くの大学が、滋賀大学モデルを参考にして、データサイエンス分野の教育を強化することを期待したい。

キーワード：ビッグデータ、課題解決型学習、アクティブラーニング、MOOC。

1. はじめに

スマートフォンが現れてから10年、情報の流れがすっかり変わってしまった。今では例えば電車の中でもほとんどの人がスマートフォンを操作しており、新聞や本を読む人は激減してしまった。人々はスマートフォンを使って常時ネットにつながっている。そして人々の行動履歴はネットに蓄積される。まさにビッグデータの時代である。このビッグデータを活かす者が競争的優位に立つ時代であり、ビッグデータから価値を引き出すための方法がデータサイエンスである。

滋賀大学は、2017年4月(平成29年)に日本初のデータサイエンス学部を開設し、新入生110名(男81, 女29)を迎え入れた。それに1年先立ち、データサイエンス教育研究センターを設置し、新学部の準備を行うとともに、データサイエンスに関する価値創造プロジェクト研究などの活動を推進してきた。このセンターは新学部設置後も存続し社会連携や研究活動を進めて行く。

さらに、滋賀大学は、2016(平成28)年12月21日に、文部科学省の数理及びデータサイエンス教育の強化に関する懇談会において、東京大学などを含め、データサイエンス教育の拠点校6つの内の1つに選定された(数理及びデータサイエンス教育の強化に関する懇談会, 2016)。この拠点校の選定は、2012(平成24)年から5年間実施された統計教育大学間連携ネットワーク

[†] 滋賀大学 データサイエンス学部：〒522-8522 滋賀県彦根市馬場一丁目1番1号

* corresponding author: shizue-izumi@biwako.shiga-u.ac.jp

(Japanese Inter-university Network for Statistical Education, JINSE)での活動の成果と関連がある(統計教育大学間連携ネットワーク, 2013, 2014, 2015). 滋賀大学は, 2016年度からJINSEに加入し, 2017年以降は統計教育連携ネットワーク(Japanese Inter-organizational Network for Statistics Education, 拡大版JINSE)の中で中心的な役割を担う.

データサイエンスの技術的基礎をなすのは, データを処理するためのデータエンジニアリング(情報学)とデータを分析するためのデータアナリシス(統計学)である. これらは工学部でも学べる理系的な手法である. ただし, これらのスキルを身につけるだけでは不十分であり, 人々の行動履歴や購買履歴などのデータに基づきビジネスや政策などの(主に文系的な)領域の課題を読み取り, 様々な分析結果を現場の意思決定に生かすために, 文系的な発想やセンスも求められる. その意味でデータサイエンスは優れて文理融合的である. ゆえに, 滋賀大学データサイエンス学部が育成したい人材は, 理系のスキルを持つとともに文系的なマインドも持つデータサイエンティストであり, これが滋賀大学の教育モデルである.

日本では, いままで統計学部や統計学科が存在しなかったこともあり, データサイエンティストの人材不足は深刻である. このことは最近の政府の文書(日本経済再生本部, 2016, 2017等)でも強く指摘されており, 実際多くの企業がデータサイエンティストを採用したい, あるいは従業員を再教育したいというニーズをもっている. 滋賀大学データサイエンス学部は, 日本初の学部であり, 今後も日本でこの分野のトップランナーとして, 多くの有能なデータサイエンティストを輩出していく.

データサイエンス教育の中で, 一番困難な部分はデータからの価値創造である. データサイエンスにおける価値創造とは, データエンジニアリングとデータアナリシスのスキルを身につけるだけでなく, 様々な領域の課題を読み取り, 分析結果を現場の意思決定に活かして課題の解決を目指すことを意味する(Maruyama et al., 2015). たとえば, マーケティングの場合, これまでに蓄積された販売データから新たな解釈を導き出し, 顧客に新たなサービスを提案することにより, 市場の拡大を図る. このようなデータサイエンスにおける価値創造は講師から学生への一方向型の講義形式で教えることはできず, 実際のデータを使ったPBL(Project-Based Learning)演習によって学生自身が成功体験を積んで行くしかない. データは企業や自治体が扱っている実際のデータであることが望ましい. しかしながら, 個人情報や企業秘密の問題があり, 実際のデータの教育利用は一筋縄では行かない. 事実, 学部に先立ち設置したデータサイエンス教育研究センターの主な活動は, 企業や自治体との連携を具体化することにあった. PBL演習の準備は現在進行中であり, カリキュラムの年次進行に沿って生のビッグデータに近いデータからの価値創造のプロセスが経験できるように演習のコンテンツを整備していく. PBL演習では, 課題を解決するために, Problem-Plan-Data-Analysis-Conclusion(PPDAC)サイクルのフェーズごとのプロセスを踏んで, 実際のデータの分析結果から, ビジネスや生活に活かすところまでの経験を積む(深澤 他, 2013).

本論文では, 滋賀大学のデータサイエンス教育の内容を概説し, 全国の大学へデータサイエンス教育のノウハウを提供する. まず2節で滋賀大学モデルの概要をまとめる. 続いて3節では, データサイエンス学部の情報系と統計系のカリキュラムについて, それぞれ述べる. 4節では, 3節で述べるカリキュラムをサポートする学習環境について述べる. 5節ではPBL演習のための国内の組織との連携の実例を幾つか紹介し, 最後に6節でまとめを行う.

2. 滋賀大学モデルの概要

2.1 データサイエンス教育研究拠点の形成

日本初のデータサイエンス教育研究拠点として, データサイエンス教育研究センターは2016

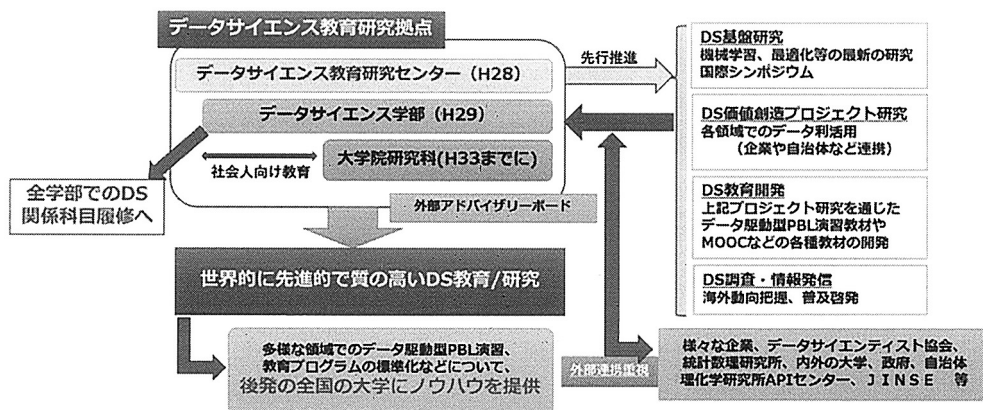


図 1. データサイエンス教育研究拠点の機能の概略。

(平成 28)年 4 月に、それに続いて、データサイエンス学部は 2017(平成 29)年 4 月に設立された。この拠点では、膨大なデータから有用な知見を引き出し、新たな価値を創造するデータサイエンスに関する先端的教育研究活動を行うとともに、企業や自治体との連携、多様な大学間連携を通じて、様々な分野における新たな価値創造、社会貢献、教育開発を行う(滋賀大学, 2017a)。専任人員は、情報系教員 8 名、統計系教員 10 名、社会調査系教員 1 名からなる。学内の教育学部や経済学部からの協力に加えて、非常勤の特別招聘教員やクロスアポイントメントの教員 9 名が活動を支援する。

この拠点は、大きく分けて次の 4 つの機能を有する(図 1)。ここでは、データサイエンスを DS と略記する。

(1) DS 基盤研究

データサイエンスの基盤となる機械学習、最適化、人工知能などの最先端の学術的な活動を進める。2016(平成 28)年度には、教員 4 名が学術団体や社会団体の賞を受賞した。

(2) DS 価値創造プロジェクト研究

企業や自治体との連携活動により、各領域でのデータに基づく課題解決を通して価値創造を図る。2017(平成 29)年 6 月時点では、20 の団体との連携を締結した。

(3) DS 教育開発

日本初のデータサイエンス教育プログラムを開発するとともに、データ駆動型 PBL 演習教材の開発や、インターネットを使用する教育サービスである大規模公開オンライン講座(Massive Open Online Courses, MOOC)などの開発を行う。2016(平成 28)年度に開発した MOOC のオンライン講座「高校生のためのデータサイエンス入門」を 2017(平成 29)年 8 月に開講した(滋賀大学, 2017c)。この講座は、基本的なデータ分析に触れる機会を高校生に与えることを目的としている。講座では、インターネット上で容易に閲覧、取得できる公的データを使って、代表値の計算やグラフの作成、また二変量データや時系列データを分析するための方法を説明する。この教材は、AO 入試においても活用する。

そして、データサイエンス研究科の早期設置を目指して、大学院経済研究科博士前期課程のプロフェッショナル・コースに「データサイエンスモデル」を 2017(平成 29)年度に開設した。データサイエンス学部の教員が担当する科目より 4 科目の履修単位が揃うと、このデータサイエンスモデルの修了が認定される。初年度には、大学院生 11 名がデータサイエンスモデルの

科目を履修した。データサイエンス研究科では、データサイエンス学部のカリキュラムをさらに体系化し、データ構造の奥深くへと潜りこみ、そこにあるデータの価値を探究するデータサイエンティストを育成していく(和泉・Gould, 2010)。

(4) DS 調査・情報発信

データサイエンスに関する海外動向や人材育成法の調査、また学問としてのデータサイエンスの動向を調査し、調査結果を発信する。2016(平成28)年度には、アメリカの統計教育やシンガポールのデータサイエンス教育について調査を行った。加えて、滋賀大学国際シンポジウム「Workshop on Undergraduate Education of Data Science」での議論を通して、DS教育研究の国際的ネットワークを形成した。2017(平成29)年度には、滋賀大学データサイエンス学部開設記念ワークショップを統計数理研究所と共同開催し、高等教育におけるデータサイエンス教育の新たな展開について議論し、PBL演習事例を紹介した。さらに、総務省統計研究研修所との連携活動の一環として、2016(平成28)年度から自治体の職員を対象にした「データサイエンスセミナー」を、2017(平成29)年度から教育関係者向けセミナーを共催する。

上記の機能のもと、データサイエンス専門の学部や大学院等で教育実践し、洗練された日本初の本格的なデータサイエンス教育を推進していく。

2.2 カリキュラムの概要

データサイエンティスト育成に焦点を合わせたカリキュラムを整えた(図2)。まず、学生は、入学後の2年間で、データエンジニアリング系科目(情報科学概論、計算機利用基礎など)やデータアナリシス系科目(基礎データ分析、線形代数など)を通して基礎知識を学習する。また、データ解析科目(基礎情報活用演習A及びB)で、統計的データ分析を行うための様々なツールの使い方を学習する。加えて、価値創造基礎科目(データサイエンス実践論A及びB、実践データ概論A及びBなど)を通して、様々な領域の専門知識に触れながら、それぞれの分野特有のデータ、解析手法、解釈の方法の特徴を理解し、多種多様なデータに対する姿勢を身につけ、データサイエンスの具体的な活用方法や価値創造のノウハウを学ぶ。それらの基礎を身につけた後に、3、4年次では、DS専門科目(バイズ理論、品質管理など)や価値創造応用科目(ファイナンス論、ファイナンス演習など)を通して、データサイエンスの分析方法を適用する実践的訓練の場において、様々な応用手法、技術とともに様々な領域の知識を身につける。さらに、学生は、1年次からデータ駆動型PBL演習(課題解決型学習)での経験を段階的に積み上げ、問題解決の様々な側面に触れる。そして、連携先の企業から提供を受けた実際のデータを用いた課題に取り組む。このような演習を4年間積み上げることにより、分析のための専門知識(文字TやIIの横棒に相当)だけではなく、領域の幅広い知識(文字TやIIの縦棒に相当)も身につけ、それらを課題解決に活かすための実践経験を積んだ「逆T型人材」あるいは「逆II型人材」を育成する。

加えて、情報技術者試験(基本情報技術者試験、応用技術者試験)、統計検定(準1級、2級)、品質管理検定(2級)の合格を目指したカリキュラムを提供している。そして、所定の単位を習得することにより、社会調査士の資格を取得できる。

2.3 初年次教育

学生は、入学後の2年間においてデータサイエンス基礎科目を履修する。たとえば、1年次前期のプレゼンテーション論では、Microsoft OfficeやBI(ビジネスインテリジェンス)ツールを使った効果的なプレゼンテーションについて理解し、各々が分かりやすいプレゼンテーションを行えるようになることを目標とする。授業では、解説に加えて、学生自身が持参したノー

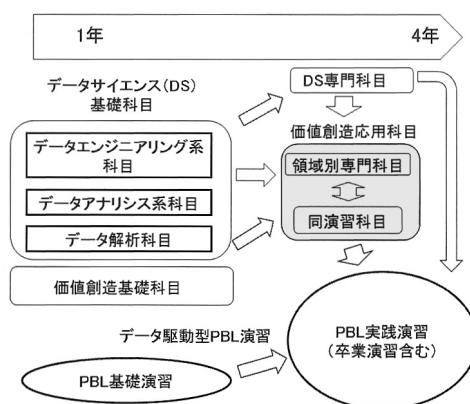


図 2. データサイエンス学部カリキュラムの模式図。

ト PC を用いて、添え状、レジュメ、統計グラフ、スライドを作成する演習を行う。一方、前期の基礎データ分析では、記述統計で使われる手法を理解し、自ら使えるようになることを目標とする。この授業では、フリップ式のアクティブラーニングを取り入れ、日本統計学会の協力のもとに作成された MOOC のオンライン大学講座「統計学 I：データ分析の基礎」を視聴して授業の予習を行い(竹村 他, 2016)、教室での授業に備える。これに続き、後期の統計学要論では、データの特性値から、そのデータの背後にある母集団の性質を統計的に推測しようとする推測統計学の基礎を理解し、自ら使えるようになることを目標とする。この授業でも、日本統計学会と日本計量生物学会の協力のもとに作成された MOOC のオンライン大学講座「統計学 II：推測統計の方法」を視聴して授業の予習を行い(竹村 他, 2015; 櫻井 他, 2015)、教室での授業に備える。加えて、教室内では実際のデータに基づく課題にグループで取り組み、課題解決のプロセスを体験する(和泉, 2017)。欧米のハイブリッド(ブレンディッド)式のアクティブラーニングを通して、統計的なコンセプトを自身の言葉で表現できるようになる(和泉 他, 2016; 深澤 他, 2017)。データサイエンス基礎科目の詳細については、3 節にて触れる。

2.4 入学者選抜の方法

入学者選抜においては、文・理両面の総合的な基礎学力を大学入試センター試験の結果によって確認しつつ、次の 3 つのタイプの選抜方法により、それぞれが重視する学力項目を評価して選抜する。

(1) 一般選抜・前期日程

英語と数学の基礎的な知識・技能を重視する教科型とする。数学の範囲は、数学 I、数学 A [全範囲]、数学 II、数学 B[(数列)と(ベクトル)]までの知識で受験可能なものとする。したがって、共通問題は、数学 I、数学 A[全範囲]、数学 II、数学 B[(数列)と(ベクトル)]の範囲とする。一方、選択問題は、①共通問題と同じ出題範囲、②数学 B[(確率分布と統計的な推測)]、③数学 III の 3 つの範囲から出題する。

(2) 一般選抜・後期日程

英語と、課題解決に向けた思考力・表現力を重視する総合・合科目型とする。総合問題では、社会や日常生活での課題をとりあげた図や表を含む文章を素材に、表やグラフを読み取り、それらを用いてデータを分析し、分かったことをまとめ、その解釈について議論する能力を問う(櫻井 他, 2011; 和泉 他, 2012; 深澤 他, 2015)。総合問題については、数学 I[(データの分析)]

を中心に幅広く出題する。

(3) アドミッション・オフィス(AO)入試

課題発見・解決力、その基礎となる思考力・判断力・表現力、学ぶ意欲、そして価値創造への主体的姿勢等を含めた「実践的な学力」の総合評価を重視する。AO入試の場合、3段階の選考を行う。

2017(平成29)年度入学者選抜試験は次の形式を用いた。第1次選考では、出願書類(調査書、志望理由書、データ分析レポート)と小論文の結果を併せて、総合的に選考した。後期日程の総合問題と同じく、小論文ではデータや図表を含む資料を読み、設問に対する論述を行った。資料や設問には、英文や数式を含むことがある。第2次選考では、データ分析レポートの内容に関するプレゼンテーションを含む個人面接により選考した。最終選考では、大学入試センター試験で受験を要する教科・科目(5教科7科目又は6教科7科目)の合計得点により最終合格者を決めた。

2018(平成30)年度入学者選抜試験からは、【AO入試I】データサイエンス講座受講型、【AO入試II】オンライン講座受講型、【AO入試III】実績評価型の3タイプに分かれる。【AO入試I】では、滋賀大学彦根キャンパスで開講するデータサイエンス講座の受講を必須とし、講座を受講後、講義内容に関する課題レポートを提出する。【AO入試II】は、滋賀大学データサイエンス教育研究センターが開発したMOOCのオンライン講座「高校生のためのデータサイエンス入門」の視聴を必須とし、出願時に、教材内容に関する課題レポートを提出する。【AO入試III】は、全国規模で開催されるデータ分析やプログラミングに関するコンペティション等への参加経験者を対象とし、出願時に、コンペティション等での実績報告レポートを提出する。どのタイプにおいても、統計質保証推進協会主催 統計検定(3級以上)、日本規格協会主催 品質管理検定(3級以上)、情報処理推進機構主催情報処理技術者試験(各種)、全国商業高等学校協会主催 情報処理検定試験(各部門第1級)の取得資格を第1次選考の評価に含める。2017(平成29)年度入学者選抜試験と同様に、最終選考では、大学入試センター試験の教科・科目の合計得点により最終合格者を決める。なお、詳細は、2018(平成30)年度入学者選抜要綱や大学ウェブページにて公表する。

2017(平成29)年度入学者選抜試験の結果、前期日程では志願者数202名の内91名が合格、後期日程では志願者数186名の内27名が合格、AO入試では志願者数17名の内3名が合格であった。

2.5 インターンシップ、卒業後の進路

現在、多数のデータサイエンティストが必要とされているため、データサイエンス学部の卒業後は、次のような業種への進路が期待される。

- 情報通信、情報技術開発
- 官公庁、地方自治体
- メーカーでの統計分析、情報解析部門
- 銀行、証券会社、保険会社等の金融機関
- 製薬、医療機関、臨床検査機関
- 経営コンサルティング、アクチュアリー
- 大学院進学(データサイエンス、その他の諸分野)

企業、自治体等と連携協定を結び、学生のインターンシップや卒業後の受け入れ口の開拓を進めている。

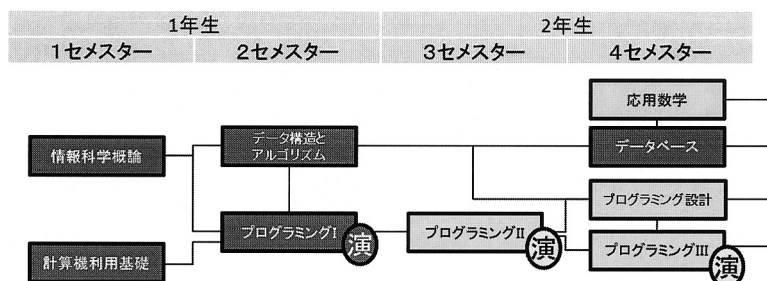


図 3. データエンジニアリング系科目のカリキュラムツリー 1, 2 年次.

3. データサイエンス学部のカリキュラム

カリキュラム・ポリシーにしたがって(滋賀大学, 2017b), データサイエンティストを育成していく。4 年間で統計学と情報学の専門的知識を習得し, データサイエンティストとしての基礎的な能力であるデータ利活用力やコミュニケーション力を養成する。

中高生が思い描く将来についての意識調査 2017(ソニー生命保険株式会社, 2017)では, IT エンジニア・プログラマーが男子中高生の考える理想の職業とされる。その背景には AI(Artificial Intelligence)や IoT(Internet of Things)に代表される次世代の技術, 人型ロボット, AR(Augmented Reality, 拡張現実)といった夢のある技術への注目と期待がある。データサイエンティストは, AI をはじめとして IoT, ネットワークセキュリティといった最先端技術に対応するスキルを持つ職業であると考えられる。AI 技術を活用するためには最新の統計的知識, AI を適用するドメインの知識とビッグデータの処理能力も求められる。そして機械学習, ニューラルネットワーク, ディープラーニングといった AI 技術を使いこなす能力も必要である。

データエンジニアリング系では, 畳み込み・バックプロパゲーション・高速な勾配法といった AI の要素技術を使いこなすために, ベクトルや行列, テンソルといったデータ構造, 条件分岐や繰り返しといったアルゴリズムを学ぶ。また AI ツールのひとつである IBM Watson(IBM, 2017)を利用したハンズオン教育など AI ツールに触れる機会を提供する。一方, データアナリシス系では, 基本的な統計知識を学び, 3, 4 年次で「分類アルゴリズム」, 「異常検知」や「因果探索」といった AI で用いられる手法を学ぶ。

3.1 データエンジニアリング系のカリキュラム

データサイエンティストとしてデータ分析を行うために必要な情報技術, ニューラルネットワークや機械学習といった人工知能, 機械学習などの発展的科目までを学ぶ。導入教育として, 1 年次に, 計算機の基本的な操作法, プログラミングなど情報学の基礎を学ぶ。全ての学生が自分専用のノートパソコンを持ち, 講義や演習に活用することで, より実践的な教育を実現する。図 3 に, 1, 2 年次のデータエンジニアリング系科目のカリキュラムツリーを示す。背景が黒の科目は必修科目を, 灰の科目は選択科目を表す。また, 「演」のマークがついている科目は演習科目も付属することを示す。

3.1.1 1 年次のカリキュラム

1 年次では, 計算機の基本的な操作法, プログラミングなど情報学の基礎を学ぶ。プログラミングの講義と演習を合わせて行い, 知識の定着を図る。プログラミング言語は主に Python を用いる。

「情報科学概論」では情報科学の基礎概念について網羅的に学ぶとともに、システム設計やアルゴリズム、プログラミングの導入教育を行う。基礎概念としてコンピュータの歴史、ハードウェアとソフトウェアの基礎、ネットワークとセキュリティ、人工知能、ニューラルネットワークといった第4次産業革命を担う技術を学ぶ。

「計算機利用基礎」では計算機システムやネットワークを安全に利用するための基礎的な概念・知識を学び、情報リテラシー能力を修得する。学習・研究における書類作成、文献調査のために Microsoft Office を用いて基礎的な資料作成方法を学ぶ。また、プログラミング言語と環境の基礎を学ぶ。

「プログラミング I」では Python の学習を通してプログラミングの基礎を身につける。コンピュータ上での情報の取り扱われ方や、動作の仕組み、プログラミング言語とは何か、どのように利用するのかなどに関する基礎をまず学ぶ。その後、プログラミング言語の具体的な文法や、変数や関数といった基礎的な知識を学ぶ。また、制御構文などを用いて、解決すべき問題に対するプログラムの作成方法を、演習を通して身につける。

「データ構造とアルゴリズム」ではいくつかの基本的なアルゴリズムとデータ構造、計算の量を客観的に評価する方法を学ぶ。スタックやキューなどのデータ構造に加えて、行列やベクトル、パーセプトロンや勾配法といった機械学習で必要となるアルゴリズムを理解するための基礎力を養う。

3.1.2 2年次のカリキュラム

2年次では、プログラミングの基礎をふまえ、応用的なプログラミング技術を学ぶ。また、新たにプログラミング言語 Java を学ぶ。さらに、データベースやプログラムの設計技術、情報学に必要な数学を学ぶ。

「プログラミング設計」では基礎的なプログラム設計の原理・技法を学ぶことができる。データ構造に関する基本概念を含め、プログラム設計表現の中で良く使われる構造図やフローチャートなどの知識をベースに、具体的な設計の表現・仕組みの基本的な知識を学ぶ。その上で、初歩的なアルゴリズム分析の意味や役割を理解し、プログラムの基本要素を組み合わせ、プログラムを組み立てて行く方法を身につける。

「プログラミング II」ではプログラミングの基礎をふまえ、応用的なプログラミング技術を学ぶことができる。主にプログラミング言語は Java を用いる。「プログラミング III」では、Java や Python を利用してシミュレーション技法やデータ解析のためのデータモデリング手法を学ぶことができる。例題として乱数アルゴリズムやモンテカルロシミュレーション、パーセプトロン、SVM(Support Vector Machine, サポートベクトルマシン)分類器といったプログラムを作成する。

「データベース」ではデータベースの知識とプログラミングを融合して情報処理システムを開発するスキルを会得する。情報処理システムの開発を通してスキルの獲得のみならず論理的な分析、設計法を学び、論理的思考力を涵養する。また AI ツールである IBM Watson (IBM, 2017) を利用した課題解決型の講義を行う。各自がシステム化すべき問題を見だし、分析手法を選択し、データ分析を行う。

3.1.3 3, 4年次の主な講義と演習

3, 4年次では、情報学の発展的な内容を学ぶことができる(図4)。コンピュータ・ネットワークシステムの制御やセキュリティ、人工知能やモバイルアプリの開発など、最先端の情報科学・技術に関連する科目が開講される。

データサイエンス教育研究センターに設置された高性能の計算サーバを利用して実習を行う。実習で用いるデータも用意され、機械学習のアプリケーションに適用することができる。

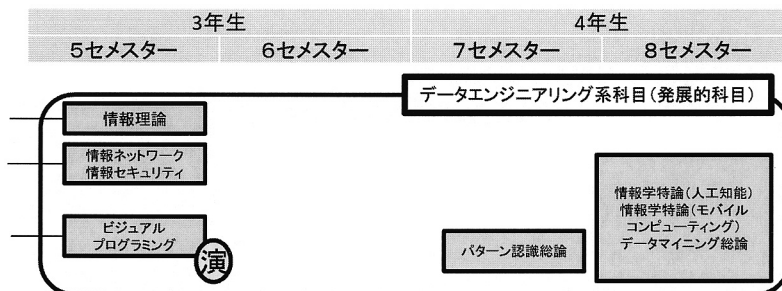


図 4. データエンジニアリング系科目のカリキュラムツリー 3, 4 年次.

畳み込みやバックプロパゲーションなど近年のディープラーニング手法、ライブラリの利用法、大量のデータの処理方法である並列・GPU プログラミングも学ぶことができる。

「情報学特論(人工知能)」では人工知能及び関連領域の多岐にわたる諸分野を概観し、後半は最近のプレイクスルーをもたらした論文を取り上げ、理論と技術を根本から学ぶことができる。基本的なアルゴリズムと統計・機械学習の初歩を学び、プログラムによる可視化方法も取り入れる。

「情報セキュリティ」では情報化社会において、適切に行動するためには、どういった使い方が危険で、どのようにすれば情報や身を守り安全に利用できるかということ学ぶことができる。ネットワークの仕組みをある程度知った上で正確な知識を身につけるために、ネットワークの基礎からセキュリティの基本的な事項までが学習内容に含まれる。

3.2 データアナリシス系科目のカリキュラム

データアナリシス系科目では、多様な構造をもつデータを、様々な分析目的に応じて分析するための、幅広い統計的分析手法に関する講義がある。これらの講義を通して、統計手法の理論だけでなく、これらをどの場面で、どのように適用すればよいか理解することを目的とする。そこで、多くのデータアナリシス系科目は、例えば「多変量解析入門」と「多変量解析」というように、本編の科目に対して入門編の科目が付属する。入門編では、各統計手法がどのような目的で使われているか、また、そこから得られる結果の意味を理解し、どのように解釈すべきかを学ぶ。つまり、数理的な内容に先立ち、手法の概要について学ぶのである。一方、本編では、各統計手法の理論的背景、導出を学び、統計スキルの向上を図る。また、講義は、座学だけでなく、計算機を用いて実際にデータ分析を行うものもある。近年ではデータ分析のためのツールやソフトウェアが数多く開発されているが、演習科目では特に Microsoft Excel, R, SAS, SPSS, JMP などのツールの使い方を学び、データ分析の経験やスキルを積む。

加えて、「標本調査法」や「社会調査法 I 及び II」といった社会調査系の専門科目がある。この知識を前提として 1 年間の社会調査実習を「社会調査実践演習」で行う。この社会調査に関する専門科目と、統計系科目のいくつかの単位を取得することで、社会調査士の資格を得ることができる。

3.2.1 1 年次のカリキュラム

1 年次のデータアナリシス系科目は、「基礎データ分析」と「統計学要論」で、データの扱い方や統計的思考の基礎を学ぶ(図 5)。まず、「基礎データ分析」では、離散データや連続データといったデータの種類やグラフの見方について学ぶ。また、クロス集計の方法や、平均値や相関係数といった記述統計量についても学習する。さらに、回帰分析や時系列データについても簡

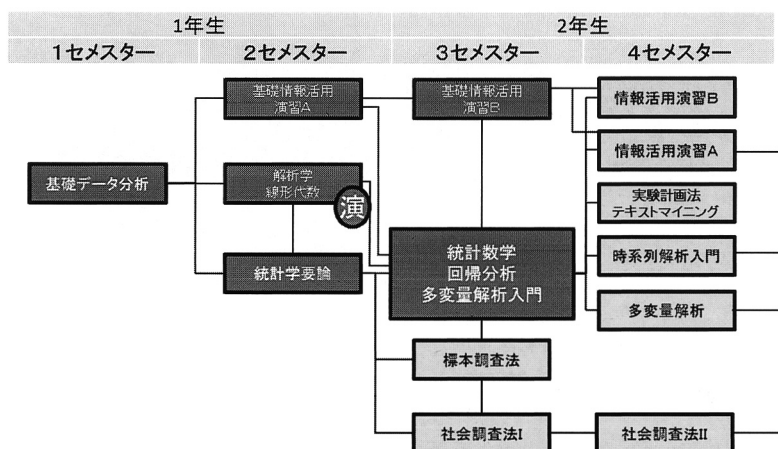


図 5. データアナリシス系科目のカリキュラムツリー 1, 2 年次.

単に触れる。

次に、「統計学要論」では、統計的思考の鍵となる母集団と標本の考え方について学ぶ。また、確率変数や確率分布の概念を学び、統計的推定や検定を理解するための下地を学習する。統計的推定や検定については、点推定や区間推定、母平均の差の検定や分割表の比率の検定など、基本的な設定における方法について学ぶ。

さらに、データ解析科目の「基礎情報活用演習 A」を通じて、Microsoft Excel を使ったデータ分析の経験も積む。Excel に用意されている統計的データ分析を行うためのツールや関数の使い方について学び、早い段階からデータ分析を行う。実際の操作としては、平均値など記述統計量の計算から始まり、統計的推定や検定、回帰分析や分散分析などを Excel 上で行い、これらの手法の理解を深める。

統計理論の理解に必要な数学的知識を「解析学」や「線形代数」で学ぶ。特にこれらの科目は、前期、後期で時間をかけて教育し、文系の学生でも身につけることができるカリキュラムになっている。「解析学」では、確率論で必須となる極限やテイラー展開、数値計算の一種であるニュートン法など、統計学に関わる内容について積極的に学ぶ。微分や積分についても、統計的推測を行う上で必要となる 2 変数関数に対する偏微分、重積分まで学ぶ。「線形代数」では、回帰分析における推定量の導出や、多変量解析に基づくデータ分析において必要となる行列演算や固有値の計算、またベクトル空間などの概念などについて学習する。

3.2.2 2 年次のカリキュラム

2 年次では、1 年次に学んだ内容の理論的背景について学ぶほか、統計的データ分析手法について学ぶ。さらに、より多様なツールやプログラミングを用いて、データ分析の経験を積む。

「統計数学」では、1 年次に学習した数学的知識を使って、離散型確率変数と連続型確率変数に関する理論についてより詳しく学ぶ。さらに、統計的理論で最も重要な要素の一つである中心極限定理や大数の法則についても学ぶ。統計的推定や検定についても、数学的背景について学習する。

「多変量解析入門」では、多変量データを分析するための手法である回帰分析や判別分析、主成分分析や因子分析、クラスタリングなどの手法の基礎を学び、それぞれの手法の役割や、得られた結果が意味するものについて学習する。これらの手法の数学的な詳細については、続く

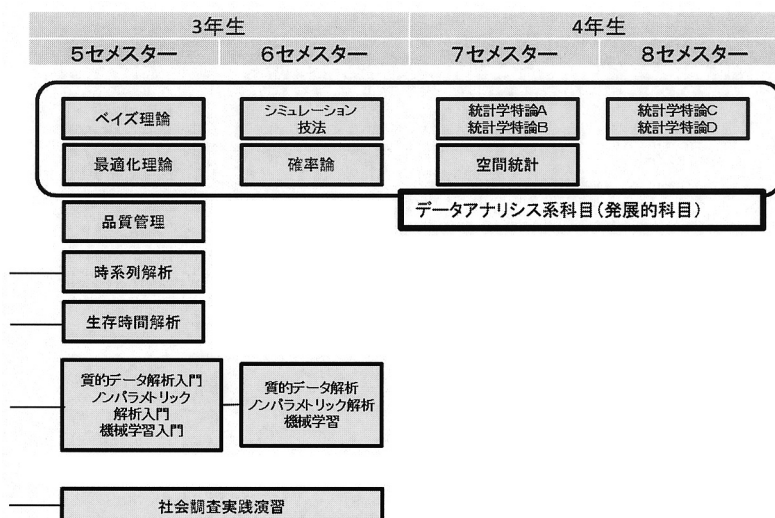


図 6. データアナリシス系科目のカリキュラムツリー 3, 4 年次.

「多変量解析」で学習することができる。また、「時系列解析入門」では、時系列データを分析するための手法である移動平均や平滑化、自己相関係数や自己回帰モデルなどの考え方について学ぶことができる。この科目についても、数学的なより詳細な内容については、3年次の「時系列解析」で学ぶことができる。

また、統計分析のツールとして開発されたソフトウェア R の利用法を「基礎情報活用演習 B」で学習し、簡単なデータ分析が R で行えるようにする。さらに、統計的データ分析のためのツールやソフトウェアとして、JMP や SPSS の利用法を「情報活用演習 A」で、SAS の利用法を「情報活用演習 B」で学ぶことができる。いずれの講義も、計算機を用いて実際に操作しながらデータ分析の方法について理解を深める。

3.2.3 3, 4 年次のカリキュラム

3年次では、より進んだ統計的思考能力や分析手法を身につけるための多様な科目が開講される(図 6)。具体的には、近年の機械学習で重要な役割を担っている「ベイズ理論」、製造業において必要不可欠な「品質管理」、医学データなどの解析で用いられる「生存時間解析」や「質的データ解析」、母集団に特定の分布を仮定せずに解析を行う「ノンパラメトリック解析」、近年の人工知能技術において中心的な要素である「機械学習」などの統計手法を学ぶことができる。また、統計学における理論的な詳細をより深く学ぶ「確率論」や、統計的推測を行う上で必要となる「最適化理論」についても学ぶことができる。これらの科目で学んだ手法や考え方を、3年次から本格的に始まる PBL 演習に活かす。

そして4年次では、さらに進んだ統計理論を「統計学特論 A, B, C, D」で学ぶことができる。内容としては、統計的推論、項目反応理論、情報量規準、計算機に基づく統計的推測の理論、空間統計などが含まれる。4年次の科目は隔年開講の予定である。

4. データサイエンス教育をサポートする学習環境

これまでに解説してきたカリキュラムをサポートするために、DS ラーニング・コモンズと教育学習支援室を用意している。まず、DS ラーニング・コモンズはデータサイエンス学部の教

育支援の中核であり、自習のための学びの場を提供している。プロジェクターやホワイトボードが2箇所設置され、ミーティングや自主ゼミに活用することが出来る。その他にも、データサイエンス関連書籍の閲覧、パソコンなど備品の貸出、そして学生生活をサポートするチューターが待機している(2017年前期では週に1回90分)。さらにDSラーニング・コモنزの管理室内には人型ロボット「Pepper」(ソフトバンクロボティクス(株))も常駐しており、プログラミング教育に活用されている。

次に、学習教育支援室は滋賀大学彦根キャンパスの施設であり、経済学部と共有している。この施設は、学生に対する学習支援、教員に対する教育支援を実施している。学生に対して、学習に必要な物品の貸し出しの他、学習スペース内のミーティングやゼミ用のブースの予約、外国語学習教材や図書の貸し出しなどを行っている。教員に対して、授業資料印刷の補助、マークシート式試験の読み取り、FD活動の支援などを行っている。

以上のような学習環境と学習教育支援体制のもとで、データサイエンス教育を行っている。将来的には、TAやSAの活用も検討しており、学生数が増えるに従って、教育支援をさらに手厚く行う予定である。

(1) 資格取得者と学業成績優秀者の表彰

資格取得者や学業成績優秀者を表彰することによって、学習のモチベーションを高める工夫を行っている。例えば、資格については、統計検定、品質管理検定、データベーススペシャリスト試験、TOEIC、税理士試験、公認会計士試験、日商簿記検定試験、証券アナリスト試験などに対して、一定の基準を満たしたものに対して報奨が与えられる。また成績については、4年間の成績が優秀な約5名を「学業成績優秀者」として卒業時に氏名を公表するとともに、成績原簿・成績証明書の特記事項欄に「成績優秀者表彰」と記入する。

(2) パソコンの環境

多種多様なビッグデータを扱うにあたり、幾つかのソフトウェアを学部単位及び全学でライセンス契約をしている。データサイエンス学部では、全学生にノートパソコンの利用を推奨しており、安全にノートパソコンを利用するためのセキュリティソフトや、様々な資料作成のためのMicrosoft Officeを提供している。また、データ分析前におけるデータの概要の把握を容易にするBIツールであるTableauやQlikSenseを利用できる環境を提供している。そして、データ分析ソフトとしては、フリーソフトであるRのみならず、JMP、SPSS、SASの利用環境を整えている。さらなる高度な分析のため、ArcGISやクラウド環境の提供も行っている。

(3) JINSE 及び拡大版 JINSE との関わり

JINSEは、2012(平成24)年度大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」に基づいて作成されたネットワークである。ネットワーク完成当初は、東京大学、大阪大学、総合研究大学院大学、青山学院大学、多摩大学、立教大学、早稲田大学、同志社大学の8大学により構成されていたが、2016年度より滋賀大学も加わった。2016年度は運営委員、カリキュラム策定委員、質保証委員として参加し、本ネットワークの策定したカリキュラムに基づくテキストの執筆にも関わった(岩崎・姫野, 2017; 統計教育大学間連携ネットワーク, 2017)。

2017年から本ネットワークは、統計教育連携ネットワーク(Japanese Inter-organizational Network for Statistics Education)と名前を変えている(統計教育連携ネットワーク, 2017)。滋賀大学は引き続きこのネットワークに参加することで、拡大版JINSEの提供する教材の利用及び、統計検定を利用した統計の学習達成度評価の分析を行っていく(田中 他, 2015)。

5. PBL 演習

データ駆動型 PBL 演習科目を 1 年次から 4 年次まで毎年実施する。これらの演習では実際のデータを利用し、様々な課題を読み取り、課題解決案を検討する。1, 2 年次(データサイエンス入門演習, データサイエンスフィールドワーク演習)では、扱いやすい課題に限定し、初歩的な分析を行い、価値創造の経験を積み重ねる。3, 4 年次のデータサイエンス実践価値創造演習 I 及び II, データサイエンス上級実践価値創造卒業演習 I 及び II はゼミ形式で実施される。情報系教員が担当するゼミでは、機械学習やディープラーニング, ネットワーク, Web プログラミング演習が行われる。統計系教員が担当するゼミでは、スパース推定, 因果推論, 漸近理論などの統計手法を駆使しつつ, データ分析演習が行われる。

データサイエンスでは, データエンジニアリング, データアナリシスに加え, 価値創造の経験を積み重ねることが重要となる。これらの演習では, 単にデータ解析の経験を積み重ねるだけではなく, 以下の段階を踏むことで, データサイエンスに関する興味, 意欲を高め, データの背景情報の重要性を知り, そのうえで価値創造の経験を重ねていく。

- 様々な分野におけるデータサイエンスの事例を聞くことで, データサイエンスへの興味・関心を高める。
- 実際の現場において, データがどのような形で収集され, どのような課題のために利用されているかを知り, 現場ごとのデータの特性を知る。
- 実際のデータを用いた分析を行い, 価値創造の経験を積む。

校内の環境だけではこれらを全て実施することはできないため, 様々な企業, 機関, 大学等の協力が不可欠である。そのため, データサイエンス学部ではこれまでにのべ 100 社以上の企業等とのコンタクトを取り, 協力を依頼してきた。そのうち, 実際に連携に至ったものは 1~2 割程度である。継続的に十分な教育環境を整えるためには, 企業と密接な連携を続けていくことが重要となるが, 特に実績が見えてくるまでにはかなりの時間と労力を要する。以下では, その成果の一部を紹介する。

(1) データサイエンスの活用事例

データサイエンスは, 環境, 生物, 防災, 医学, 品質管理, 交通等, 様々な分野で活用されている。これらの分野の専門家を滋賀県立大学, 長浜バイオ大学, 京都大学, 滋賀医科大学, 積水化学工業株式会社, 名古屋大学から招聘教員またはクロスアポイントメントという形で招き, 各分野での事例紹介を授業内で行っている。また, 企業でのデータサイエンスの事例については, 日本アイ・ビー・エム株式会社, 株式会社野村総合研究所, データサイエンティスト協会, NPO 法人ビュー・コミュニケーションズ, 第一生命ホールディング株式会社の方々をゲストスピーカーとして招き, 実践的, 先進的なデータサイエンスの事例紹介, ハンズオンなどを行っている。

(2) 工場見学

データの背景の重要性を知るため, 1 年次に工場見学を実施している。毎年, 4~5 企業の工場を 20~40 人ずつの学生が見学し, データ活用に関する事例を学ぶ。2017 年度は地元企業 4 社による協力のもと工場見学を実施した。

(3) 実際のデータを利用した PBL 演習

実際のデータを教育用として使用するには様々なハードルがある。教育用としてのデータ提供となると, 企業側のメリットが少なく, データ提供の際, 個人情報保護のための匿名化処理や秘密保持契約の締結などの手続きも必要なため, 企業から協力が得られるのはまれである。このような中, 株式会社オプト, 株式会社アイディーズから教育システムの開発・提供を受け,

NPO 法人ビュー・コミュニケーションズからは複数のデータを提供されている。

株式会社オプトは、データ解析コンテスト DeepAnalytics を教育利用できるように、新たに DeepAnalytics for Education を開発した。このシステムを利用すると、トレーニングデータに基づき、テストデータの予測を行い、その予測結果をこのシステムに投稿することで、即座に予測精度を評価し、予測スコア及び投稿者のランキングが表示される。学生間でスコアを競いながら予測精度向上を目指すことで、様々な工夫を行い、実践的なデータ分析が行われる。このシステムでは、様々なデータサンプルが用意されており、数多くのデータ予測コンテストを実施することが可能である。

株式会社アイディーズからは生鮮食品を含む食品購買ビッグデータ i-code の提供を受け、様々な演習へ活用していく。ここでの i-code は生鮮食品・惣菜も網羅した我が国初の「商品標準化コード」であり、商品が購入された年月日、時間帯だけでなく、同時に何が購入されているのかも知ることができ、様々な価値創造を行うことが期待される。

6. 終わりに

冒頭にも述べたように、日本におけるデータサイエンティストの人材不足は深刻である、このことは最近の政府の文書でも強く指摘されており、例えば『日本再興戦略2016』のIV.3.(2)項では「高等教育での数理情報教育を強化するため、トップレベルデータサイエンティストなどを育成する学部・大学院の整備促進するとともに、全学的な数学教育の強化(標準カリキュラム策定)等を実施する」としている。滋賀大学のデータサイエンス学部の設立はこのような方針に沿ったものである。また滋賀大学がデータサイエンス教育の6拠点校に選定されたのは、以上で説明してきたデータサイエンス学部の体系的なデータサイエンス教育の内容が評価されたためである。データサイエンス分野の日本での遅れを取り戻すべく、多くの大学が滋賀大学を一つのモデルとしてデータサイエンス分野の教育を強化することを期待したい。

謝 辞

本論文で紹介したデータサイエンス教育に関して様々なご支援・ご教示を頂きました統計数理研究所の田村義保先生に深く感謝申し上げます。なお、本論文の一部は、国立大学改革強化推進補助金(総合支援型)事業推進費、統計数理研究所・共同研究利用(重点型研究)課題番号29-共研-4306の補助を受けた。

参 考 文 献

- 深澤弘美, 櫻井尚子, 和泉志津恵 (2013). 学校教育における問題解決型統計教育とその評価, 「統計教育実践研究 第5巻」, 統計数理研究所共同研究リポート, No.293, 62-65.
- 深澤弘美, 櫻井尚子, 和泉志津恵 (2015). 大学入試センター試験と海外の出題傾向, 「統計教育実践研究 第7巻」, 統計数理研究所共同研究リポート, No.335, 69-72.
- 深澤弘美, 和泉志津恵, 櫻井尚子 (2017). 統計教育における評価指標の作成と試行—高校, 大学の教養レベル—, 「統計教育実践研究 第9巻」, 統計数理研究所共同研究レポート, No.379, 9-14.
- IBM (2017). IBM Watson, <https://www.ibm.com/watson/jp-ja/>.
- 岩崎学, 姫野哲人 (2017). 『スタンダード 統計学基礎』, 培風館, 東京.
- 和泉志津恵 (2017). 公的ビッグデータを用いた課題解決型学習(PBL)演習の試行, 「統計教育実践研究 第9巻」, 統計数理研究所共同研究レポート, No.379, 97-102.
- 和泉志津恵, Gould, Robert (2010). 大学・大学院における統計教育の国際比較, 「統計教育実践研究 第

- 2 巻], 統計数理研究所共同研究レポート, No.243, 43-46.
- 和泉志津恵, 深澤弘美, 櫻井尚子 (2012). 高等学校での数学 I「データ分析」に関するアセスメントの提案, 「統計教育実践研究 第 4 巻」, 統計数理研究所共同研究レポート, No.272, 30-33.
- 和泉志津恵, 櫻井尚子, 深澤弘美 (2016). 大学の統計教育でのインタラクティブな授業のデザインとその評価方法, 「統計教育実践研究 第 8 巻」, 統計数理研究所共同研究レポート, No.362, 5-10.
- Maruyama, H., Kamiya, N., Higuchi, T. and Takemura, A. (2015). Developing data analytics skills in Japan: Status and challenge, *Journal of Japan Industrial Management Association*, **65**, 334-339.
- 日本経済再生本部 (2016). 日本再興戦略 2016—第 4 次産業革命に向けて—, 2016 年 6 月 2 日閣議決定, http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/2016_zentaihombun.pdf.
- 日本経済再生本部 (2017). 未来投資戦略 2017—Society 5.0 の実現に向けた改革, 2017 年 6 月 9 日閣議決定, <http://www.kantei.go.jp/jp/singi/keizaisaisei/>.
- 櫻井尚子, 和泉志津恵, 深澤弘美 (2011). 国内外で実施された統計学に関する各種試験問題の考察, 「統計教育実践研究 第 3 巻」, 統計数理研究所共同研究レポート, No.260, 59-64.
- 櫻井尚子, 和泉志津恵, 深澤弘美 (2015). MOOC での certification 問題について, 「統計教育実践研究 第 7 巻」, 統計数理研究所共同研究レポート, No.335, 73-78.
- 滋賀大学 (2017a). Data Science View, Shiga University, Vol.1, 国立大学法人滋賀大学データサイエンス教育研究センター, 滋賀.
- 滋賀大学 (2017b). 滋賀大学データサイエンス学部カリキュラム・ポリシー, http://www.shiga-u.ac.jp/information/info_public-info/public-education/public-education_hikone/.
- 滋賀大学 (2017c). オンライン講座「高校生のためのデータサイエンス入門」, 国立大学法人滋賀大学データサイエンス教育研究センター, 滋賀, https://lms.gacco.org/courses/course-v1:gacco+pt010+2017_08/about.
- ソニー生命保険株式会社 (2017). 中高生が思い描く将来についての意識調査 2017, http://www.sonylife.co.jp/company/news/29/nr_170425.html.
- 数理及びデータサイエンス教育の強化に関する懇談会 (2016). 「数理及びデータサイエンスに係る教育強化」の拠点校の選定について, http://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/1380792.htm.
- 竹村彰通, 椎名洋, 和泉志津恵, 松田安昌, 佐藤俊哉 (2015). 『統計学 II: 推測統計の方法 オフィシャルスタディノート』, 日本統計協会, 東京.
- 竹村彰通, 酒折文武, 中山厚穂, 下川敏雄 (2016). 『統計学 I: データ分析の基礎 オフィシャルスタディノート 改訂版』, 日本統計協会, 東京.
- 田中豊, 中西寛子, 姫野哲人, 酒折文武, 山本義郎 (2015). 『改訂版日本統計学会公式認定統計検定 2 級対応統計学基礎』, 東京図書, 東京.
- 統計教育大学間連携ネットワーク (2013). 平成 24 年度(2012)総合報告書, <http://jinse.jp/pdf/24souhou.pdf>.
- 統計教育大学間連携ネットワーク (2014). 統計学の各分野における教育課程編成上の参照基準, <http://jinse.jp/pdf/Sansyou20140515.pdf>.
- 統計教育大学間連携ネットワーク (2015). 平成 25 年度(2013)活動報告書, <http://jinse.jp/pdf/25rep.pdf>.
- 統計教育大学間連携ネットワーク (監修)(2017). 『現代統計学』, 美添泰人, 竹村彰通, 宿久洋(編集), 日本評論社, 東京.
- 統計教育連携ネットワーク (2017). 統計教育連携ネットワーク規約, http://www.qajss.org/jinse_kiyaku_20170808.pdf.

Shiga University Model of Data Science Education

Akimichi Takemura, Shizue Izumi, Kunihiko Saito,
Tetsuto Himeno, Hidetoshi Matsui and Heiwa Date

Faculty of Data Science, Shiga University

One method for extracting value from big data is data science. The technical foundations of data science are data engineering (information science) and data analysis (statistics). Shiga University established a faculty of data science in April 2017, in order to cultivate data scientists with liberal arts educations along with skill in the physical sciences. In this article, we explain the content of data science education at Shiga University. First, we summarize the Shiga University Model. Then, we describe the curriculum for information science and the statistics of data science, respectively, and we describe a learning environment to support these curricula. Furthermore, we introduce examples of cooperation with domestic organizations for Project-Based Learning (PBL) practices. We anticipate that many universities in Japan will strengthen education in the field of data science using the Shiga University model as a reference.