

数理・データサイエンス・AI 教育の紹介

弘前大学の数理・データサイエンス・AI教育プログラム

～グループワークとデータサイエンスを融合した課題解決学習と岩木健康増進プロジェクトでのデータの利活用～

弘前大学
数理データサイエンス教育センター長
理工学研究科教授

守 真太郎

弘前大学
大学院医学研究科附属健康・医療データ
サイエンス研究センター長
医学研究科教授

玉田 嘉紀

弘前大学
数理データサイエンス教育センター
教育推進機構・教育戦略室助教

じょ きょうてつ
徐 貺哲



(左から 守、玉田、徐)

1. はじめに

本学では令和2年度から教養教育科目として数理データサイエンス（MDS）教育を試行的に実施し、令和4年度からリテラシーレベルプログラムを全学1年次必修科目、応用基礎プログラムは2年次以降に履修する選択科目として実施しています。本稿では、まず、MDS教育プログラムについて説明し、その後、リテラシーレベルプラスの認定理由とされた「特色ある課題解決学習」と「岩木健康増進プロジェクトのデータ利活用とデータサイエンス教育での利用方針」について説明したいと思います。数理・データサイエンス・AI教育プログラムを準備・実施し、リテラシーレベ

ルの認定やプラスでの選定を目指している担当者の方の参考となれば幸いです。

2. データサイエンス教育プログラム

図1に、本学のMDS教育プログラムの概要を示します。プログラムでは、新入生は前期に「データサイエンス基礎」、後期に「地域学ゼミナール」を必修科目として履修します。図では緑の枠で囲まれている2科目4単位で数理・データサイエンスリテラシープログラムは構成されています。

応用基礎プログラムは1年前期の必修科目「データサイエンス基礎」を軸に選択科目の「データサイエンス数学」、「データサイエンス発展I、II」

を加えた4科目8単位で構成されます。図では青文字で示しています。応用基礎レベルのプログラムではAI・データサイエンスを学ぶための数学を学ぶことが必須です。「統計学」の講義は多くの学部・学科で開講されているのですが、微積・線形代数はそうではありません。「データサイエンス数学」は専門教育で微積や線形代数を扱う授業のない学生のために開講しています。理工学部や教育学部で

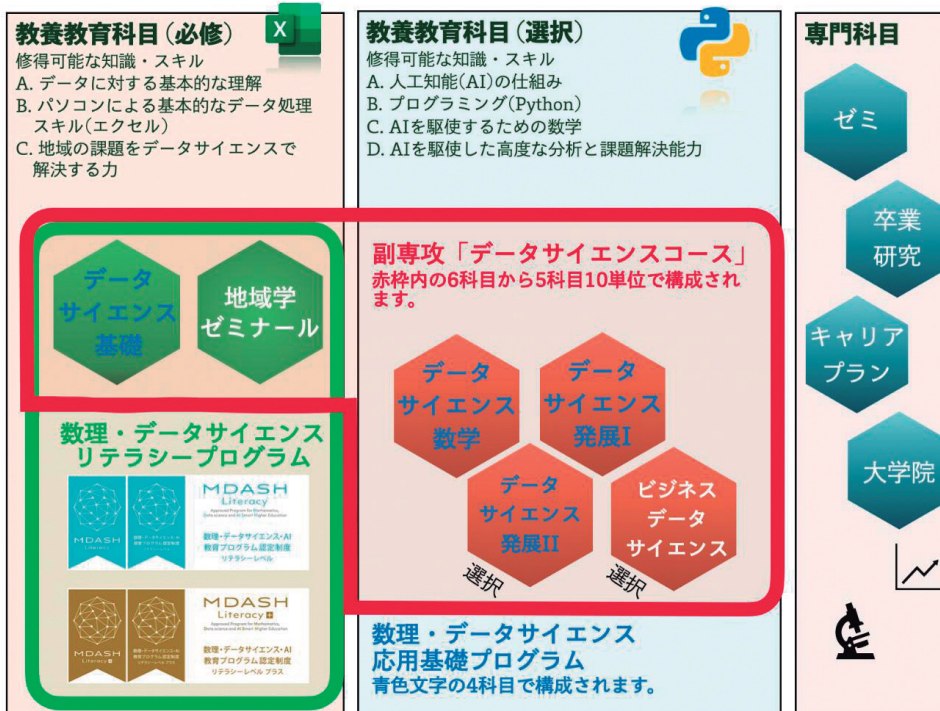


図1 データサイエンス教育プログラムの概要

は専門科目として微積・線形代数を扱う講義は開講されていますので、応用基礎プログラムでは専門教育科目でも代替可能としています。また、「データサイエンス発展I、II」ではPythonを使用して回帰分析からディープラーニング、自然言語処理までの内容を学びます。

最後の赤枠で囲んで示しているのは副専攻プログラム「データサイエンスコース」です。本学ではリテラシープログラムは必修としていますが、応用基礎プログラムを構成する科目は選択科目となっています。「データサイエンスコース」は応用基礎プログラムの履修を推進するために用意した履修プログラムです。説明の仕方にもよるとは思いますが、学生にとってMDASH認定プログラムと副専攻プログラムのどちらが分かりやすいかと考えたとき、「主専攻にプラスして副専攻でデータサイエンスを学ぶ」のほうが理解しやすいと考えたためです。

3. プログラムの実施体制

約1,400名の新入生が最初に受講するのが「データサイエンス基礎」です。令和4年度は16クラスが開講され、学生は概ね所属学科の先生の講義を受講します。学生20名から25名に対してTAを1名配置し、Excelを用いたデータ分析演習で学生のフォローを行っています。1年次後期に履修する「地域学ゼミナール」は、学部横断型のクラス編成とし、グループワークの授業を行います。令和4年度は54名の教員と同数のTAで18クラスが開講されました。教員1名TA1名がユニットとなり25名程度の学生の指導にあたります。

2年次以降は選択科目として「データサイエンス発展I、II」をそれぞれ前期と後期に履修します。履修者数に制限は設けず、履修希望者は全員履修可能です。時間割の都合で履修できない学生には長期休暇中のオンデマンド授業も実施しています。

担当は本学のデータサイエンス教育プログラムの実施・改善・点検を行う数理・データサイエンス教育センター¹⁾所属の教員です。センターには各学部から「データサイエンス基礎」を担当する教員と、教育推進機構に所属するデータサイエンスを専門とする教員、および理工学部でプログラミングの授業を担当する教員が兼任で参加してい

ます。「データサイエンス発展I、II」は教育推進機構のデータサイエンスを専門とする教員が主に担当しています。20名から25名の学生に対して1名のTAが付きまゝ。「データサイエンス数学」、「ビジネスデータサイエンス」もデータサイエンスを専門とする教員が担当しています。

リテラシープログラムを構成する「データサイエンス基礎」、「地域学ゼミナール」はデータ分析でExcelを用いています。本学ではPC必携化とマイクロソフトとの包括契約により、学生がOfficeをいつでも利用できます。

応用基礎プログラムを構成する「データサイエンス発展I、II」では(株)Signateのe-Learning教材Signate Cloudを採用し、シラバスに合わせた教材を選んでコースを設定しています。学生は各自のPCから教材にアクセスして履修します。「データサイエンス発展II」では最終課題に取り組みます。応用基礎プログラム履修後も機械学習を用いてデータ分析できる環境として、本学の情報基盤センターがJupyter HUBを用いたプログラミング環境を提供しています。最終課題はJupyter HUBで取り組むこととしています。

4. プログラムの各科目の内容紹介

(1)「データサイエンス基礎」

MDS教育プログラムの根幹となる科目です。初回のガイダンス後、①データサイエンスの応用事例、②データ利活用プロジェクトの進め方、③データ利活用とリスク、人工知能の倫理的問題、について第2から第5回にオンデマンド形式で学びます。オンデマンド期間中は学生は教室に来て受講し、TAに質問することもできますし、在宅受講も可能です。各回の最後に小テストを用意し、第6回冒頭の間テストで学生の修得状況を確認します。中間テストの正解率は約90%となっており、オンデマンド授業でも学生はしっかり勉強して試験の対策をしていることが分かります。

第6回後半から第15回はPCを用いたデータ分析演習です。第6回の講義の後半に、授業の後半の目標が重回帰分析を使って問題解決フレームワークのPPDACサイクルを経験することと説明します。その上で、第7回以降、①データマネジメント、②質的データの分析、③量的データの分析、



図2 「データサイエンス基礎」テキスト

④推測統計の基礎、を学びます。講義は、毎回10分講義+30分演習のセットを2セット行う形式とし、10分の講義で学んだ内容をデータとExcelを用いて確認します。

教科書はセンターの教員で分担執筆したものを採用しています（図2）。また、講義スライド、講義動画、演習課題、演習解説動画はMoodleに配置し、学生が予習・復習しやすく、前半（第2回から5回）の講義と同様、オンデマンド授業にも対応できる体制としています。

第14回は総復習回とし、データサイエンスを用いた問題解決フレームワークであるPPDACサイクルを復習し、医療費の予測課題に重回帰分析で取り組みます。第15回には第16回の期末試験で分析するデータを配布します。担当教員の判断で、期末試験の模擬問題の演習を行うことも、独自の教材でデータ分析課題演習を行うことも可能です。

オンデマンド回の確認の小テスト、第6回の中間テスト、第7回から14回の課題、第16回の期末試験はMoodleの小テスト、または、マイクロソフトのFormsを用いて実施し、自動で採点と集計を行っています。採点結果はセンターで管理し、担当教員の負担の軽減を行っています。

（2）「地域学ゼミナール」

「データサイエンス基礎」で学んだデータサイエンスと1年前期必修科目のスタディスキル科目

「基礎ゼミナール」で学ぶブレインストーミング、KJ法などのグループワーク手法を用いて地域課題の解決に取り組めます。リテラシープログラムの総仕上げ的な位置づけの科目です。詳細は次の第5節で説明します。

（3）「データサイエンス発展Ⅰ、Ⅱ」

「データサイエンス基礎」で学んだことをプログラミングと機械学習による分析手法を修得してパワーアップし、専攻分野でのデータサイエンス利活用につなげるための科目です。

「データサイエンス基礎」では、Excelを用いて重回帰分析できることを目標としています。いわゆるノーコードの分析となりますが、巨大・複雑なデータを分析するにはExcelでは限界があり、プログラミングが不可欠となります。データ分析のプログラミング言語としてはR、Pythonの二つの選択肢がありましたが、機械学習ではPythonが主流なこと、学生が企業に就職後にデータ分析で用いる言語はPythonが主流であることからPythonとしています。

「データサイエンス発展Ⅰ」では前半の6回でPythonプログラミングを学び、データマネジメント、データ可視化を扱ったのち、重回帰分析を汎化能力の観点で学びなおします。

「データサイエンス発展Ⅱ」では、ニューラルネットワーク、機械学習の手法、ディープラーニング、自然言語処理を学び、最後の5回でデータ分析演習に取り組みます。Kaggleで行われた過去のデータサイエンスコンペを教材としています。

5. 地域課題解決のグループワーク演習

地域学ゼミナール（以下「地域学ゼミ」と略す）では、学部等の垣根を超えて、自分とは異なる興味・関心や背景を持つ他者と協力して、チームで問題解決に取り組めます。1年次前期にグループワークの方法論をスタディスキル科目「基礎ゼミナール」で、データサイエンスを「データサイエンス基礎」で学び、両科目の総仕上げを「地域学ゼミ」で行います。青森県の実際の問題をテーマとし、その解決策をグループワークとデータサイエンス分析能力を駆使して提案し、最後にプレゼンを行います。

異なる考え方や感じ方を持つ異質な他者とチームを組んで問題解決に取り組んだ経験は、学生が社会で働き、良い市民として生きていく上で重要な意味を持ちます。また、現実の問題を解決するには、複眼的に問題を捉えたり、解決策の実行がもたらす複合的な影響を解明し、その成果や副作用を多角的に評価することが欠かせません。

演習は3つの部分で構成されています。第一部では、基本的なスキルを習得するための座学を行います。ブレインストーミングやPPDACサイクル（図3）を復習し、グループワークでの問題解決方法を理解します。

第二部では、表1にリストアップされた9つのテーマからテーマを選択し、以降、PPDACサイクル（図3）を実践していきます。第二部では、データの収集・集計や可視化、回帰分析などのデ

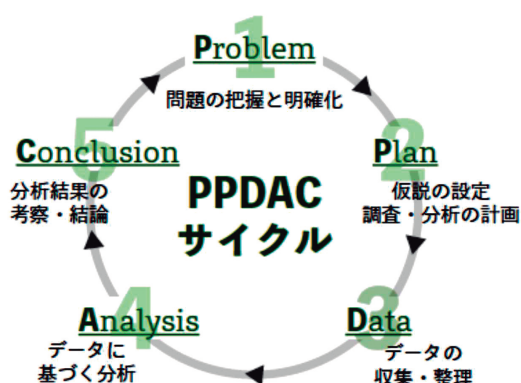


図3 PPDACサイクル
(出典「データサイエンス基礎」弘前大学出版会)

表1 地域課題のテーマのリスト

1. 弘南鉄道大鰐線を存続させるには？
2. 共に生きるあおもり ―子育ての男女共同参画実現に向けて取り組むべきことは？―
3. りんごを海外で売りまくれ！！ ―青森県のりんご輸出量を倍増させよ―
4. 若者の政治参加を促すには？
5. 白神山地魅力再発見 ―白神山地観光客数の減少を食い止める―
6. 身近なところから短命県返上 ―弘大生の生活習慣病リスクゼロに挑戦―
7. 買い物弱者を救え！
8. オラこんな村イヤだ ―急募：若者を青森に惹きつける方法―
9. 学生目線の商品開発 ―地域資源を生かした特産品を生み出す―

ータサイエンスの手法を活用することに重点を置き、「データサイエンス基礎」で学んだ知識を実践する機会としています。また、第二部終了後の中間発表では分析結果を報告する機会も設けています。第三部では、選択したテーマに対して解決策を考え、最終発表会で発表し学生同士で評価し合います。

地域学ゼミの最大の特徴は、学部横断の取り組みです。そのため、授業内容も文理融合を重視しています。地域課題をデータサイエンスの問題解決フレームワークであるPPDACサイクルに落とし込むため、文献調査やエビデンスに基づく問題の理解も重視しています。また、チームで課題解決に取り組むため、各自の得意分野を活かすこともできます。例えば、Aさんが文献調査、Bさんがデータ収集、Cさんがデータ分析、Dさんがスライド作成や統括を担当するなど、役割分担が可能です。各自が得意分野を活かして協力し、課題解決に最も有効な方策を提案する能力を身につけることができます。

担当教員は第一部と第二部の座学終了後は、演習の進行を管理するのみで、最終発表までグループワークにはタッチしません。TAと協力し、学生からの質問に答えるだけです。学生（1年生）に任せてしまって、まともな解決案にたどり着くのかと思われるかもしれませんが、データ分析と深い考察に基づくプレゼン内容に驚かされます。以下、2022年度に徐が担当したクラスのなかから興味深い発表を行った3つのグループの成果を紹介します。

（1）グループ1の研究発表

テーマ：「若者政治参加に関する研究」

このグループは若者が政治参加に関心が低いという問題を取り上げ、その原因と改善案について検討しました。

文献調査により、若者の政治参加が消極的になっている原因を探りました。先行研究の知見を参考にし、自ら収集したデータを用いて重回帰分析を行いました。その結果、投票に行かない人の特徴として、「政治と生活は関係ない」という認識が強いことがわかりました。一方、投票に行く人の特徴としては、「政治が良くないので変えたい」

という考えを持っている人が多いことがわかりました。

中間発表の分析結果を踏まえて、投票に行かない人を3つのグループに分類し、それぞれ異なる解決策を提案しました（図4）。

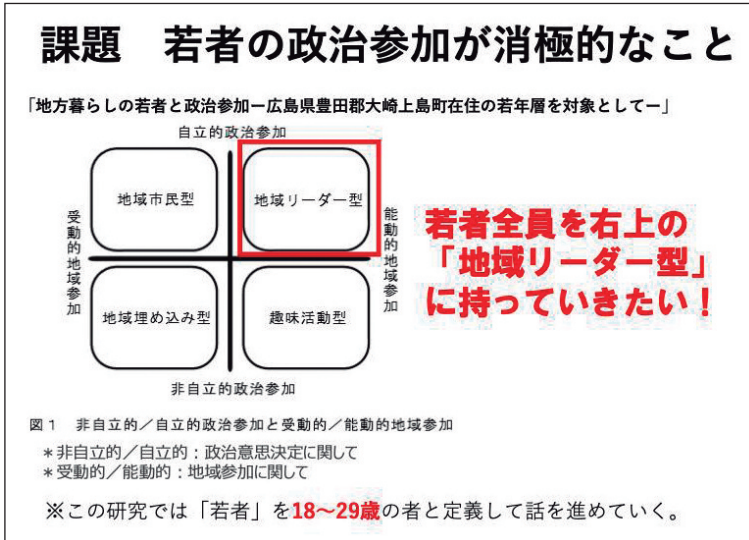


図4 グループ1の最終発表スライド（抜粋）

（2）グループ2の研究発表

テーマ：「身近なところから短命県返上」

青森県は、都道府県別の平均寿命ランキングでほぼ最下位です。このグループは青森県の短命県問題について、その原因と改善案について検討しました。

中間発表では、先行研究を参考に過度なアルコール摂取による影響が問題ではないかと仮説を立て、重回帰分析を行いました（図5）。さらに、長寿県として有名な長野県のデータとも比較しま

した。その結果、青森県民は低アルコール度数のお酒を大量に摂取しているのに対し、長野県民は高アルコール度数のお酒を少量摂取していることが明らかになりました。

中間発表の結果を受け、他の生活習慣にも問題がある可能性について分析しました。その結果、喫煙や肥満などの問題も非常に顕著であることが明らかになりました。そこで、彼らが考えた解決策は、健康な飲食を促進するために管理栄養士の採用を増やすことです。健康な飲食習慣の改善により、県民の生活習慣病が改善され、寿命が延びるだけでなく、雇用の改善にもつながるとしました。

（3）グループ3の研究発表

テーマ：「道路の改善から地域活性化へ」

青森県では若者の県外流出の問題があります。グループ3では交通の利便性と若者の定住率の影響の観点で問題解決を検討しました。

中間発表では、公共交通機関の利便性向上は難しく、道路の改善で交通事故が減少し、暮らしやすさの向上と地域の活性化につながることを示されました。一方、車を持たない人は検討課題として残りました。

中間発表の結果に基づき、公共交通機関の現状を再分析し、交通機関をどのように改善すれば地域活性化につなげることができるかについて、2つの解決策を提案しました。一つは、若者にとって新鮮味のある「レトロ」な車体デザインへの変更です。車体デザインを変更し、交通機関そのもの

を活性化させることで、新たな景観づくりや、Instagramなどで写真映えする要素を求める観光客の増加が期待できます。もう一つは、本学の通学にも利用できる100円バスの増設です。地域の生活をより便利にすることで、定住者の増加を促進することを目指しています。

6. データサイエンス教育での岩木健康増進プロジェクトのデータ活用

岩木健康増進プロジェクトは2005年よ

2019年（令和元年）～2021年（令和3年）平均			
酒類	金	酒類	金
0 全国	44,076	全国	5,541
1 青森市	57,280	秋田市	10,092
2 秋田市	57,253	新潟市	9,251
3 札幌市	54,794	仙台市	8,937
4 東京都港区	53,790	富山市	8,097
5 盛岡市	53,764	福島市	7,948
6 新潟市	52,899	盛岡市	7,867
7 富山市	50,587	金沢市	7,145
8 福島市	50,302	さいたま市	6,819
9 山形市	49,023	長野市	6,671
10 相模原市	48,735	松江市	6,656
11 横濱市	48,376	水戸市	6,654
12 さいたま市	47,727	佐賀市	6,578
13 仙台市	47,614	宇都宮市	6,482
14 広島市	47,602	福井市	6,471
15 金沢市	47,325	青森市	6,458
16 千葉市	47,034	京都市	6,348
17 大阪市	47,014	千代市	6,300
18 川崎市	46,930	前橋市	6,285
19 高崎市	46,095	東京都港区	6,158

評価指標と要因

・要因
青森県は三大生活習慣病の人の数が全国の中でも上位
→それらの原因のひとつ：飲酒
→青森県は酒類消費量全国1位

・評価指標
アルコール入り飲料のアルコール含有量
年間の全酒購入金額
→これらを長野県と青森県で比較

図5 グループ2の中間発表のスライド（抜粋）

り弘前市岩木地区住民の生活習慣予防と健康の維持・増進、寿命の延伸を目指して本学大学院医学研究科が中心となって始めたプロジェクトです。様々な取組みの中でとりわけ健診事業は、住民の健康ビッグデータを長期に亘って収集し続けている大規模な健診プロジェクトとなっていて現在19年目を迎えています。毎年6月頃に集中的に健診を行い、10日間で約1,000名分の健康データを収集しています。

本健診の特徴は、3000項目にも及ぶ膨大な健診データを収集している点で、通常の健診で計測される身長や体重などに加え、ゲノムデータや腸内・口腔細菌叢、膨大なアンケートによる社会的背景や生活習慣の調査データが含まれています。医学研究科内の多くの基礎講座、臨床講座が参加しており、それぞれ独自の視点で健診項目を追加しています。医学部学生も参加しており研修の場としても役立ててもらっています。また医学研究科だけでなく理工学研究科の講座も健診に参加しています。本健診は、健康データ収集のための「プラットフォーム」となっていることも特徴となっており、多くの企業がこのプラットフォームを用いた独自データの収集を行なっています。現在およそ20社が健診に参画しており、共同研究講座を設置し共同で様々な項目を測定、収集しています。つまりデータがさらなるデータと呼ぶ状況を作り上げています。このような大規模な健康調査データは世界でもここ本学にしか存在しません。健診結果は住民にお返しし健康増進に役立てていただいているほか、市役所とも連携し政策立案や市としての住民の健康増進活動の一環として利用してもらっています。また各臨床および基礎講座などでの研究教育、参画企業での研究に活用されているほか、超多項目ビッグデータであること活かした健康AI研究にも利活用されています。本プロジェクトは、このように産官学民4者が連携しその全者がメリットを享受できるプロジェクトとなっています。

データサイエンスでは「データ利用」の側面のみが語られがちですが、岩木健康増進プロジェクトでは、研究デザインから始まり、健診現場でのデータ取得も直接行い、さらに血液などの試料は、分析・計測を経てデータ化されます。またアンケ

ートデータは欠損や整合性の確認など膨大なデータの事後整備を経て、ようやく実際に計算機で解析可能なデータになり、データ解析が行われます。つまりデータサイエンスの全ての工程・側面がこのプロジェクトに詰まっていると言えます。

本プロジェクトではこのように自ら大規模なデータ計測をしていますので、健診会場でのデータ取得プロセスの管理、集めたデータを効率よく解析するためのデータベース開発、さらにはデータの管理・運営も必要になります。これにはデータ解析技術だけでなく情報技術全般が必要になります。それらを行うための専門組織として、2023年4月より医学研究科附属健康・医療データサイエンス研究センター^[2]を立ち上げました。また、この健康ビッグデータの解析をするためにはインフラが必要で、特にゲノムデータはプライバシーに関わる秘匿性の高いデータであることからゲノム解析専用のスーパーコンピュータなどを導入・運営しています。さらに、これらの膨大な健康ビッグデータを解析するためのセミナーや、これらのインフラを活用したデータサイエンス教育も実施しています。

健康ビッグデータの解析により、様々な疾患の予兆を捉えるなどの未病領域での研究に活用されていますが、データサイエンス教育が全学で開始したことで、今後はデータに基づいて分析を行うことが身近になり、利活用が広がることを期待しています。

7. おわりに

本稿では、本学の数理・データサイエンス教育の概要とリテラシーレベルプラスの認定理由とされた二つの点について紹介しました。本学では、本学で開発した教材は無償で提供しています。ご興味があれば、数理・データサイエンス教育センターまでお問合せください。

関連URL

- [1] 弘前大学数理データサイエンス教育センター
<https://gkm.hirosaki-u.ac.jp/cmds/>
- [2] 弘前大学大学院医学研究科附属健康・医療データサイエンス研究センター
<https://ytlab.jp/rcohmds/index.html>