

「政治学特殊講義（社会問題のデータ分析入門）」実践報告 —人文社会科学部におけるデータ分析教育—

■ 小川 寛貴（高知大学人文社会科学部）

キーワード：データ分析，数理・データサイエンス・AI教育

1 背景

本稿は2020年度より人文社会科学部の専門科目として開講している「政治学特殊講義（社会問題のデータ分析入門）」の実践報告である。実践報告に先立ち、本節では開講の背景について簡単に説明しておく。

2019年6月11日に統合イノベーション戦略推進会議によって決定された「AI戦略2019」では、「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」（統合イノベーション戦略推進会議2019：12）することが具体目標の1つに掲げられた¹。

これに伴い、2021年には大学等での数理・データサイエンス・AI教育を推進するために「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）」が創設された（文部科学省高等教育局2021）。認定制度は2021年2月26日から公募が開始されており、認定にかかる審査は外部有識者により構成される審査委員会によって行われる。審査の結果を踏まえ、文部科学大臣がプログラム認定を行う仕組みである。

具体的な教育プログラムのモデルとしては、2020年

4月に数理・データサイエンス教育強化拠点コンソーシアムによって「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～データ思考の涵養～」が発表された。モデルカリキュラムは「導入（社会におけるデータ・AI利活用）」「基礎（データリテラシー）」「心得（データ・AI利活用における留意事項）」「選択（オプション）」の4つの要素から構成されている（数理・データサイエンス教育強化拠点コンソーシアム2020：7）。前述の認定制度においては、このうち「導入」「基礎」「心得」が審査項目に含まれている（文部科学省高等教育局2021：4）。2021年には、モデルカリキュラムのうち「選択」以外の項目に対応したテキスト（北川・竹村編2021）も公刊された。

本学の動きとしては、共通教育において、2021年度より「高知大学リテラシーレベル数理・データサイエンス教育プログラム」が展開されている。このプログラムでは、所定の区分から必要単位数を修得した学生について本学の数理・データサイエンス（リテラシーレベル）の修了者として認定される。

このような全国のおよび全学的な数理・データサイエンス・AI教育の展開の流れの中で、2020年度よりデータ分析に焦点を当てた科目として、「政治学特殊講義（社会問題のデータ分析入門）」を開講した。本授業はおおまかにいえばデータ分析の基礎的な技能を習得することを目的としている。前述のモデルカリキュラムや認定制度との関係でいうならば「基礎」に重き

¹ 2021年6月11日には統合イノベーション戦略推進会議により「AI戦略2021」が決定されたが、この具体目標自体には変更はない（統合イノベーション戦略推進会議2021：15）。

を置きつつ、「導入」「心得」も適宜フォローするような内容である²。本稿では、以上の経緯で開講された「政治学特殊講義（社会問題のデータ分析入門）」について、その実践報告を行い今後の展開を検討する。

本稿の構成は以下のとおりである。第2節では授業の設計について、到達目標と難易度の設定および授業で用いる統計ソフトの選定について説明する。第3節では授業実施について、詳細な実施方法と成績評価について述べる。第4節では、授業実施にあたって工夫した点を4つ紹介する。第5節では、第15回の授業で実施した簡単な授業アンケートの結果から授業を振り返る。最後に第6節では、今後の展開と課題について検討する。

2 授業の設計

2-1 到達目標と難易度の設定

本授業を設置した狙いは、データ分析への抵抗感を極力排し、可能な限り多くの学生にデータ分析の基礎を理解してもらうことにある。受講生が分析方法として使うか否かにかかわらず、様々な分野で活用されるようになったデータ分析について理解しておくことには、人文・社会科学分野の学習に加え現代社会を生きる上でも意味があるだろう。しかしながら、データ分析と聞けば一定の拒否反応があることも否めない。また、分析で用いる統計ソフトの操作に抵抗感があることも否めない。そこで本授業では、あまり受講のハードルを上げないような形式と難易度で、なるべく多くの学生にデータ分析の基礎を学んでもらおうと考えた。

以上の背景から、本授業で想定した受講生は、これまでにデータ分析に触れたことはないが、データ分析に漠然とした興味をもつ学生である。それゆえ、履修にあたって前提知識は求めず、特定の科目履修要件も定めなかった。加えて、データ分析の各手法の説明に当たり数学的説明は必要最低限に留め、基礎的な内容

はさらいつつも筆者が専門とする政治学分野の標準テキストの1つである浅野・矢内（2013, 2018）を少々易化した難易度とした。

本授業では、データ分析（最終到達点は重回帰分析）を実行してその成果を文書（期末レポート）で報告できるようにすることを、最終的な到達目標とした。卒業論文等でデータ分析を利用する場合の最低限の手法の習得が本授業の狙いである。この最終的な到達目標を踏まえて、シラバスに示した授業の到達目標は、①データ分析の原理について説明できる、②データ分析を行うにあたり必要な操作を実行できる、③自分が興味のあるテーマについてデータを収集し分析することができる、④データ分析の成果を文書の形で他者に報告（説明）できる、の4つとした。

全15回の構成は表1の通りである。第1回ではデータ分析の意義や活用例を説明する。第2回ではデータ分析工程の全体像を示し、後述する統計ソフトの導入等の準備を行う。第3回から第10回まではデータ分析の原理の学習と実践を行う回である。内容としては記述統計から入り、推測統計、仮説検定と歩みを進めて最終的に重回帰分析に到達する構成である。第11回以降は各自が定めたテーマに基づく分析を実行する回である。

表1 授業の構成

回数	内容
第1回	イントロダクション
第2回	データ分析の第一歩
第3回	記述統計Ⅰ（統計量）
第4回	記述統計Ⅱ（可視化）
第5回	推測統計
第6回	仮説検定の考え方（平均値の検定）
第7回	2つの要素の関係性を調べる（クロス表分析・相関分析）
第8回	単回帰分析
第9回	重回帰分析
第10回	発展的な分析
第11回	分析に向けた準備
第12回	データの集め方
第13回	分析の実践Ⅰ
第14回	分析の実践Ⅱ
第15回	授業のまとめ

各回ではデータ分析の原理に関する学習と共に、統計ソフトを利用した分析の実習を行う。詳細な授業形態と受講の流れについては第3節で詳しく紹介する。

² データ分析の原理に関する内容の一部は「選択」あるいはリテラシーレベルではなく応用基礎レベル（数理・データサイエンス教育強化拠点コンソーシアム2021を参照）にも該当する。

2-2 利用する統計ソフト

実習を伴うデータ分析の授業において重要なポイントの1つは利用ソフトの選定である。想定される受講生像を前提として、利用ソフトのコストや操作性をよく吟味する必要がある。統計ソフトは数多く存在するが、筆者が少なくとも1回は触れたことのある統計ソフトから例を挙げると、Stata (Light Stone 社³)、EViews (Light Stone 社⁴)、SPSS Statistics (IBM 社⁵)、SAS (SAS Institute 社⁶)、R (フリーソフト⁷)、HAD (フリーソフト⁸) などがある。このように複数ある選択肢からどのソフトを採用するかによって、授業の事前準備や、学生の費用負担および操作上の負担は大きく変わってくる。今回は以下の2点を重視して利用ソフトを選択した。

1点目は、フリーソフトかどうかである。統計ソフトの購入コストは非常に高いため、一つの授業のために受講生に有料ソフトを購入してもらうことは難しい。学内のパソコンに統計ソフトがインストールされていればコスト面の心配はないが、残念ながら人文社会科学部棟内には20人規模の授業で利用可能なPCルームはない。仮に統計ソフトを備えたPCルームがあったとしても、対面授業の展開が困難な近況では活用しにくい。また、統計ソフトは可能な限り受講生のパソコンに入っていることが望ましい。なぜなら、統計ソフトの利用スキルの習得にあたっては繰り返しの操作が欠かせないが、学内のパソコンでは学生が任意のタイミングで繰り返し操作して分析することはできない。一方で、自分のパソコンにソフトが入っていれば、いつでも統計ソフトに触れることができる。以上の点から、フリーソフトのRあるいはHADなどが候補ということになる。また、Microsoft Excelも本学で

は学生の費用負担がないという意味で候補に入れることができる。

2点目は、操作性である。ここでいう操作性とは、受講生にとって簡単に操作できるか、を問うている。この意味でRはハードルが高い。Rには様々な周辺ツールが存在しており、例えばRStudio⁹を利用すれば分析作業の利便性は上がり、R Markdown¹⁰を併用すればレポートなどの成果物の作成も容易である。しかしながら、初めてデータ分析に触れる受講生を想定し、プログラミング等の知識を前提としないならば、Rでの授業展開は難しい。なぜなら、分析実行のために必要なコードの書き方や、様々な場面で生じるエラーへの対応が必要だからである。浅野・中村(2018)のような詳細なエラー対応の記述があるテキストがあるにせよ、エラー対応やコードを教えることに多くの時間が取られる可能性がある¹¹。本授業は比較的内容が多く、あまり時間の余裕がないので、この点は大きな不安要素になる。また、対面授業であればまだしも、オンライン授業の可能性がある中では、個別のエラー対応は困難と判断せざるを得ない。

学生に親しみがある、という意味ではExcelは有力な選択肢である。Excelは表計算のソフトであるが、分析用のアドインを入れることで様々な分析を簡単に行うことができる。しかしながら、Excelでの分析には様々な制約があり、操作性や出力の面でもデータ分析用に作られている統計ソフトと比較して課題があることは否めない¹²。

そこで今回利用したソフトが、社会心理学者の清水裕士によって開発されたフリーソフトのHAD¹³である。HADはExcelベースで操作できるため、操作面での心配はあまりない。また、対応可能な分析方法も

³ <https://www.lightstone.co.jp/statAIndex.html> (2021年9月1日最終アクセス)

⁴ <https://www.lightstone.co.jp/eviews/> (2021年9月1日最終アクセス)

⁵ <https://www.ibm.com/jp-ja/analytics/spss-statistics-software> (2021年9月1日最終アクセス)

⁶ https://www.sas.com/ja_jp/solutions/analytics.html (2021年9月1日最終アクセス)

⁷ <https://www.r-project.org/> (2021年9月1日最終アクセス)

⁸ <https://norimune.net/had> (2021年9月1日最終アクセス)

⁹ <https://www.rstudio.com/products/rstudio/> (2021年9月1日最終アクセス)

¹⁰ <https://rmarkdown.rstudio.com/> (2021年9月1日最終アクセス)

¹¹ コードの煩雑さを回避するためにはR Commanderを利用する手もあるが、後述するHADの操作性との比較で今回は利用しなかった。

¹² Excel統計 (<https://bellcurve.jp/ex/>) というソフトもあるが、フリーソフトではない。

¹³ HADの詳細な説明は製作者によるウェブサイト<<https://norimune.net/had>>を参照。

非常に多様であり、因子分析や構造方程式モデル、マルチレベル分析なども実施可能である。フリーソフトで学生に親しみのある Excel ベースであること、卒業研究等でデータ分析を行う際の選択肢も多いこと、などを踏まえて本授業では HAD を利用することにした。

3 授業の実施

3-1 授業実施の概要

本節では、授業の実施方法について述べる。本授業は履修開始年次を2年生として、火曜日2時限に開講されている。2020年度からの開講であるため、講義形態としては基本的にオンラインで実施された。具体的には、2020年度は全ての回が高知大学 moodle を用いた非同期型で行われ、2021年度については第5回までは対面、第6回以降は moodle を用いた非同期型で実施された。本授業はデータ分析の実習を含み、授業後半では受講者個別のデータを用いた分析を行うが、1人の教員で個別指導およびフィードバック可能な人数には限りがある。そのため、2020年度は18名、2021年度は24名と受講人数を絞ることとした。いずれの年度も予定受講者数を10名程度超える履修登録があったため、社会科学コース2年生を優先した抽選によって受講生を決定した。なお、受講者数が制限されることと、抽選時の優先順位については事前にシラバスに記しておいた。

3-2 授業実施の詳細

次に、授業実施の詳細について説明する。以下では、本授業の大半で行われた moodle を用いた非同期型での授業実施方法について説明する。本授業の受講の流れは大きく分けて3つの段階に分けられる。

第1段階として、受講生は事前にアップロードされたテキストを読むことが求められる。このテキストは、浅野・矢内(2013, 2018)および小宮・布井(2018)を参考に作成した。テキストのアップロードは、2020年度は授業日、2021年度は授業日の1週間前に行った。

第2段階として、受講生は各20～30分の授業動画を3本視聴する。そのうち2本はその回の分析方法に関する説明動画(以下、説明編動画)である。説明編動

画の視聴にあたっては、スライドの PDF ファイルとキーワード等の穴埋めをするためのレジメ(word ファイル)を提供している。残りの1本は HAD の画面を映し出した実習用の動画(以下、操作編動画)であり、その回で学んだ分析方法を HAD で行うための手順を解説している。操作編動画で使うデータは moodle にアップロードされており、受講生は動画を視聴しながら手元のデータで同じ操作を再現してソフト操作のスキルを習得する。

第3段階として、受講生は各回で課された課題を提出する。基本的な課題内容は、提供されたデータについて、説明編動画で扱われた手法を利用して操作編動画で説明された手順に従って分析し、その結果を報告するというものである。各回の課題の解答例は、1週間後に次回の授業コンテンツと共に moodle にアップロードされる。また、課題に対する詳細な個別フィードバックは2020年度は第4回から、2021年度はオンライン講義に移行した第6回から行った。フィードバックでは、得点と共に課題の出来に対するコメントやアドバイスを受講生に返される。

なお、2020年度は操作編動画で同じ操作を再現するためのデータと、課題用データのファイルを変えていた。すなわち、2020年度は2つのデータファイルを提供していた。しかし、この方法では受講生のファイル管理が煩雑になるおそれがあるため、2021年度は同じデータファイルで提供し、操作編動画と課題で利用する変数を変えることとした。

3-3 成績評価

成績評価の内訳は、各回の課題を65%(到達目標①②③の評価)、期末レポートを35%(到達目標④の評価)とした。成績評価の65%を占める課題は、授業内で計13回課されている。課題の提出状況については、2020年度は受講生18名中16名が全13回提出し、1名が1回のみ未提出、1名が3回未提出であった。2021年度は24名中20名が全13回提出し、2名が1回のみ未提出、2名が2回未提出であった。いずれの年度においても途中で受講をやめ課題を提出しなくなる受講生はいな

かった。また、提出された課題について趣旨を外れるものはいずれの年度においても1つもなく、受講生側が十分に授業内容と課題内容を理解して取り組めていたと考えられる。

期末レポートについては2020年度、2021年度ともに全員が提出している。期末レポートの課題は各自が第11回授業で設定した複数のテーマと仮説のうち、教員側でデータが用意できたテーマと仮説について分析を行うというものである。ここで設定するテーマについては、なるべくデータ分析に親しんでほしいという狙いから、原則自由に設定してもらった。期末レポートで課された分析は、記述統計表の作成、従属変数の箱ひげ図の作成、相関分析の実施と解釈、重回帰分析の実施と解釈、の4つである。また、分析レポートの要件として、テーマ設定の背景、設定した仮説とその仮説が導出された理由、データの出典と分析で利用する変数の定義の説明、分析結果の提示と解釈、結論、のそれぞれについて記述することが求められる。なお、期末レポート執筆にむけた分析レポートの書き方については第14回授業で説明している。

期末レポート全体の出来をみると、分析部分については、記述統計表の作成、従属変数の箱ひげ図の作成、相関分析の実施と解釈、の3点はほぼ全員が問題なくできていた。重回帰分析の実施と解釈については、有意確率（p値）の取り違いによる帰無仮説の棄却判断ミスと変数の単位の取り違いによる偏回帰係数の解釈ミスが若干名に生じていたが、こちらもほぼ全員が問題なく分析を実行して解釈を行っていた。分析レポートの要件に関しては、データの出典と変数の定義のいずれかが欠落してしまうケース少々見られたが、その他の要素には概して問題がなく、全員があらかじめ要求した分析レポートの要件を十分に満たした期末レポートを執筆できていた。

最終的な成績評価の分布は表2の通りである。表2には2020年度と2021年度のそれぞれについて、各評定に該当する受講者数と当該年度の受講者数に占める割合が示されている。ここまでの課題及び期末レポートの振り返りから分かるように、2020年度、2021年度共

に受講生はモチベーション高く講義に参加し、課題および期末レポートのパフォーマンスも予定していた目標をはるかに上回るものであった。その結果として、いずれの年度も「秀」評価の者が多くなっている。

表2 成績評価の割合

	2020		2021	
秀	13	72.2%	16	66.7%
優	1	5.6%	6	25.0%
良	4	22.2%	2	8.3%
可	0	0.0%	0	0.0%
不	0	0.0%	0	0.0%

このような高いモチベーションとパフォーマンスは受講生側の意欲によるものと解しているが、本授業としても受講生のモチベーションを維持し、授業内容を理解する上での躓きをなるべく早めに解決するためのフォローを個別に行ってきた。そこで次節では、受講生のモチベーション維持と理解度向上のために、本授業ではどのような工夫をしたのかを簡単に振り返ることとする。

4 工夫した点

本節では、授業実施にあたって工夫した点を4つ挙げておく。第1に、様々な分野の分析事例を紹介した点である。データ分析を学習するにあたり、具体的な活用方法が見えてこないモチベーションが最後まで持続しない可能性がある。そこで、データ分析を活用すると具体的にどのようなことが分析できるようになるのか、なるべく具体的な研究成果を紹介するように努めた。その際、受講生の関心は多岐にわたっているため、多くの分野の事例を紹介するようにした。特に第1回では、経済学、社会学、法律学、経営学・マーケティング、心理学、文学、政治学など多様な分野における研究事例を紹介している。そうすることで、受講生が興味のあるテーマに対するデータ分析活用の具体的なイメージを持ったうえで、各回の授業に取り組めるようにした。

第2に、受講生の関心に即したデータの選択と提供を行った点である。本授業では、操作編動画と課題で実際のデータを使って分析の実習を行う。そこで用いるデータは教科書的に用意された架空のデータよりも、実際に研究等で使われるようなデータの方が受講生の関心を高められると考えた。また、データ分析に対する関心やモチベーションを維持してもらうためには、なるべく受講生の興味関心に即したデータを提供することが望ましいと考えた。そこで、第1回の課題としてデータ分析で扱ってみたいテーマについて受講生に記述して提出してもらい、その回答結果を参考にデータを収集し、授業内の分析実習で用いることとした。

また、授業後半においては各自が興味関心あるテーマについて分析し期末レポートにまとめることになるが、ここでは受講生が設定したテーマと仮説にあわせて個別にデータを提供した。具体的には、第11回授業の課題として期末レポートに向けた分析で扱いたいテーマと仮説のセットを3つ回答してもらい、その回答から分析可能なテーマと仮説について、受講生に個別にデータを提供した¹⁴。これにより、受講生全員が異なるデータを使って分析を行い、期末レポートを執筆することになる¹⁵。以上を通して、受講生がデータ分析に楽しみを見出し、なるべくモチベーションを維持できるように工夫をした。

第3に、可能な限り詳細なフィードバックを行ったことである。データ分析の学習は積み重ねが肝要なため、1つの躓きがその後の理解度に与える影響は大きい。そこで、各回の課題では受講生個々の理解度を把握するとともに、課題への回答が不十分な場合には、どこで間違いが起きていたのかが受講生に明確に分かるようにフィードバックを行った。例えば、課題において誤った数値が報告されている場合には、同じ数値が再現できるよう受講生と同じデータでの計算・分析

を繰り返して、具体的な問題点を発見し、その詳細を課題のフィードバックに付すこととした。これらのフィードバックの成果として、基本的な計算ミスや操作ミス、数値の取り違いなどが繰り返して生じるケースは少なかった。

第4に、実践的な方法論講義としての位置づけを強く意識した点である。データ分析の原理や統計ソフトの操作方法を単に教わるだけでは、受講生自身の研究においてデータ分析を実際に活用することは難しい。データ分析活用の前段として、どのような問いに対してどのような仮説を構築すればよいのか。データ分析の成果はどのように文書化して報告すればよいのか。簡単にはあるがこれらの点も扱うことで、データ分析の手法がより実践的な研究方法として受講生に身につくと考えた。そこで、受講生個人の分析を開始する授業後半においては、リサーチデザインとデータ分析の成果報告の方法も講義内容に盛り込むこととした¹⁶。具体的には、第11回と第12回では仮説の構築を含むリサーチデザインについて簡単な説明を行っている。また、第14回ではデータ分析を利用した研究レポートの書き方について説明しており、受講生はそこで示された構成に沿って期末レポートを執筆することになる。以上の内容を盛り込むことで、簡素化されたものではあるものの、受講生はデータ分析を用いた研究の一通りのプロセスを体験できるようになっている。

5 アンケートによる評価と振り返り

本節では、アンケート結果から本授業について振り返る。本授業では、授業の理解度や難易度などについて確認するため、成績評価に含めないと断った上で第15回に簡単なアンケートを無記名式で実施した。受講生の回答負荷を考慮してアンケート項目はさしあたり確認しておきたい5項目(表3参照)に絞り、質問文は本学や社会科学コースで利用されているアンケートを参考に作成した。回答率は2020年度では約83%(15/18)、2021年度では約79%(19/24)であった。

¹⁴ 本来はデータ収集もデータ分析の大事なプロセスなので受講生の手で行うことが望ましいが、本授業ではスケジュールの都合上、データ収集について詳細に扱うことが困難であったためこのような方法を採用した。

¹⁵ ただし、受講生間で難易度が変わらないように、計算が複雑で解釈が難しい変数や、専門性が高くレポートでの定義説明が困難になりそうな変数はデータに含めないようにした。

¹⁶ 例えば、浅野・矢内(2018)は全15章のうち3章をリサーチデザインの説明に充てている。

表3 アンケート項目

番号	質問文
Q1	この授業をうけて、データ分析（重回帰分析まで）の手法が身についたと思いますか。
Q2	この授業で必要とされた授業時間外の学修（課題など）は適切な分量でしたか。
Q3	授業では統計ソフトとしてExcelをベースとしたHADを利用しました。HADを実際に操作してみてもの実感として近いものを以下から選択してください。
Q4	この授業の難易度・レベルは、あなたにとってどうでしたか。
Q5	この授業を半期の間受講しての感想・コメントを自由に記入してください。

まず、データ分析の手法の理解度について、Q1の結果を表4に示した。表4には2020年度と2021年度のそれぞれについて、各選択肢を選んだ回答者数と当該年度の回答者に占める割合が示されている。表4からは、回答者のうちほぼ全員が「身についた」「どちらかといえば身についた」を選択しており、データ分析の手法が一定程度身についていると考えられる。最終成果物である期末レポートにおいても、ほぼ全員が適切に重回帰分析を実施して必要事項を報告できており、各回の課題及び期末レポートの出来とアンケート結果とのずれはないように思われる。

表4 データ分析の理解度について（Q1）

	2020		2021	
	回数	割合	回数	割合
身についた	10	66.7%	11	57.9%
どちらかといえば身についた	5	33.3%	7	36.8%
どちらともいえない	0	0.0%	0	0.0%
どちらかといえば身につかなかった	0	0.0%	1	5.3%
身につかなかった	0	0.0%	0	0.0%

次に、時間外学修の分量について、Q2の結果を表5に示した。本授業は事前の予習や事後の課題に相應の時間を要する。授業設計時には適切な負担量に調整しているつもりではあるが、受講生の認識はどうか。表5からは、2021年度に「どちらかといえば多かった」という回答があるものの、その他の回答者は全員「適切だった」を選択している。表5の結果からは、本授業が課している時間外学修の分量は適切なものであるといて差し支えないであろう。

表5 授業時間外の学修の分量について（Q2）

	2020		2021	
	回数	割合	回数	割合
多かった	0	0.0%	0	0.0%
どちらかといえば多かった	0	0.0%	2	10.5%
適切だった	15	100.0%	17	89.5%
どちらかといえば少なかった	0	0.0%	0	0.0%
少なかった	0	0.0%	0	0.0%

続いて、本授業で利用した統計ソフトHADの操作難易度について、Q3の結果を表6に示した。「難しかった」「どちらかといえば難しかった」の合計は2020年度では約47%、2021年度では約32%であり、ExcelベースといえどもHADの操作は簡単ではないようである。ソフト操作の難易度や料金等を踏まえると、HADが現行採用しうる最低限の難易度なのかもしれない。また、授業の改善点としては、ソフト操作面のフォローをより充実させる必要があるようである。

表6 HADの難易度について（Q3）

	2020		2021	
	回数	割合	回数	割合
難しかった	1	6.7%	3	15.8%
どちらかといえば難しかった	6	40.0%	3	15.8%
どちらともいえない	2	13.3%	8	42.1%
どちらかといえば簡単だった	6	40.0%	5	26.3%
簡単だった	0	0.0%	0	0.0%

授業全体の難易度については、Q4で尋ねており、その結果を表7に示した。2020年度と2021年度で共通して分かることは、簡単だったと感じる受講生は極めて少ないということである。両年共に「難しかった」「どちらかといえば難しかった」の合計が半数を超えることはなかったが、今後の展開としてさらに授業内容を増やす、より数学的な説明を増やす、といったことは難しそうである。また、第3節の成績分布と本節表4の回答結果からは受講生の理解度は高かったと思われるが、体感した難易度は相応に高かったようである。第3節でみたように、いずれの年度も受講生の参加度とモチベーションが高かったと推測されることを踏まえると、現状の難易度を最高ラインとして、今後の授業設計を考える必要があるかもしれない。

表7 授業の難易度について (Q4)

	2020		2021	
難しかった	0	0.0%	3	15.8%
どちらかといえば難しかった	7	46.7%	5	26.3%
どちらともいえない	6	40.0%	10	52.6%
どちらかといえば簡単だった	2	13.3%	1	5.3%
簡単だった	0	0.0%	0	0.0%

最後に、自由回答 (Q5) の中から受講生の実際の声をいくつかを抽出して紹介する。自由回答で最も言及が多かったのは、卒業論文を含めた今後のデータ分析の活用についてである。一例を挙げると、「今後、卒論などでデータ分析を使う場面があると思います。その際に、この授業で学んだことを生かせればいいと思います。」「卒論やレポート等でもこの授業で学んだデータ分析の手法を活用していければと思います。」「データ分析は今後とも使っていきたいと思います。」といった回答があり、本授業で学んだデータ分析の手法を今後も活用したいという意欲が示されていた。ただし、こうした卒業論文等での展開が本格化する場合には、データ分析に関する分野間あるいは教員間での認識の違いがあることが懸念される。

授業の方式については授業の大半がオンラインで実施されたことの影響が危惧されたが、「授業動画や授業テキストがとても分かりやすかったので、オンラインではあったものの授業内容が理解しやすかったです。」「解説を全くわからない自分でもわかるよう丁寧に一からしていただいたので、操作や、ないほうがよくわかりました。データを集めて下さったり、テキストが豊富であることから非常に学習しやすく、ありがたかったです。」「私は機械に本当に弱く、Excel を使って分析するのに不安はありましたが、講義資料や動画がとても分かりやすく、バックアップも手厚かったので、安心して受講でき、最後までやりきることができたと思います。」といった回答からみるに大きな問題はなかったようである。

HAD の操作難易度については、「had の操作が授業動画を何度も止めながらやらないとできなかった。それでも、説明が分かりやすかったので had の操作が完全にわからなくなることが無かったです。」と操作が

一筋縄ではいかないケースがみられた。表6の結果からも HAD の操作は想定よりも難易度が高いようなので、今後は操作マニュアルや操作編動画の充実を図り、より操作面のフォローを強化したい。

授業の難易度については「難しかったことは、帰無仮説など仮説検定のロジックの理解や有意水準・有意確率など聞きなれない言葉の理解です。」「復習しなければ、混乱することが多いと感じました。」「非常に授業が難しかったが、先生の優しいサポートがありがたかった。」といった回答があり、表7で示された結果通り体感の難易度は高かったようである。特に1番目の回答にある仮説検定のロジックやその説明で使われる「帰無仮説」「有意水準」「有意確率」は本授業で最も難易度が高い部分と思われる。もう少し時間を割いて説明する必要があるかもしれない。

理解度については、「HAD の操作の面については、少し難しいと感じることもありましたが、動画を繰り返し見たり、Teams で相談に乗ってくださったりしていただいたことで、分析の手法が身についたと思います。」「実践的に分析できたことで、講義で身につく質が良かったと考えている。」「データの読み方やデータを用いた分析の方法が身についたと感じました。」などの回答があり、分析の手法が身についたと感じている様子が窺える。

受講のモチベーションについては、「実際に使うことができれば自分の強みにもなると感じたため、受講前から最後までモチベーションを維持することができました。そのため、データ分析を学ぶメリットを教えていただいたことがモチベーションにつながっていたと思います。」と第4節でも紹介したモチベーション維持のための試みが一定程度機能していたことが窺える回答があった。また、「学んでいるこちらとしても身につけていると実感しながら学習を進められたので、課題を出して終わりのような授業と違い、とてもやる気が出ました。」と各回の内容をこなすごとに身につけていることを感じてやる気が出たという回答や、「何気に課題にコメントをつけてくれたのが毎回嬉しかったです。」とフィードバックに対する好意的

な回答もあった。これらのコメントからは、各回の課題へのフィードバックが理解度促進だけでなく、モチベーションの維持にもつながっていたことが考えられ、第4節で挙げたような工夫は今後も継続していく必要があるだろう。

最後に、データ分析に対する面白さや楽しさを見出したという回答もあった。例えば、「操作について、想像よりも簡単であり、自分の操作で分析結果を出せることが面白かったです。」「はじめは難しいと感じていたけど、仕組みが分かるようになり、データ分析の結果が読めるようになったときは楽しかったです。」「自分の予想と実際に出てくる数値が意外と異なっているものもあり面白かった。」といった回答である。やや内容が多く難易度が高い授業であったため、データ分析そのものの魅力が伝えられるかには大いに不安があったが、これらの回答を見る限りデータ分析について手法としての有意性だけでなく、面白みも感じてもらえたようである。

以上のアンケート結果と第3節の成績評価をあわせて考えると、受講生にとって授業内容や統計ソフトHADの操作は決して簡単でなかった中でも設定した目標を大きく超えるパフォーマンスを発揮したといえる。自由回答の記述からは、理解の促進やモチベーション維持のためには第3節で示した授業実施方法や第4節で示した工夫がある程度機能していたといえる。特に個別対応や詳細なフィードバックが重要な役割を果たしていたと思われるが、これには受講可能人数とのトレードオフの問題がある。すなわち、詳細なフィードバックなどのフォローを充実させるほど、大人数での授業実施は困難になるということである。次節の後半でも触れるが、この点は本授業の大きな課題である。

6 今後の展望と課題

本稿の締めくくりとして、今後の展望と課題について検討する。成績評価(第3節)やアンケート結果(第5節)からすれば、多くの受講生はシラバスに示した4つの到達目標(第2節)について当初の目標をはる

かに上回る水準で達成できていると思われる。その意味で、データ分析の初心者を対象としてデータ分析の基礎を身につけてもらう、という本授業の狙いは十分に果たしていると考えられる。

授業の内容や難易度については、第5節のアンケート結果を踏まえると、より多くの内容を盛り込む、難易度を高める、といった試みは難しそうである。また、統計ソフトについてはHADから変更する理由は現在のところ存在しない。以上の点から、単独で完結させるデータ分析の授業としては、現状で実行可能な範囲では十分な教育ができていていると思われる。

もちろん、開講の背景となった「数理・データサイエンス・AI教育」の展開としては、データ分析部分だけで見ても単独科目ゆえの限界がある。本授業の内容はかなり欲張ったものになっている一方で、前提知識を課さないが故に基本的な内容から授業を始めている。そのため、データ分析の各手法の原理の説明や分析の実習に十分な時間が割けていない可能性がある。さらに、あくまで基礎的な内容なので、授業で触れることができていない要素も少なくない。当然ながら1つの科目ですべてを扱うことはできないので、データ分析に焦点を当てた「数理・データサイエンス・AI教育」の展開としては、複数の科目を連動させるような方向性が考えられる。

例えば、前提科目として統計学の授業を受講して基礎的な内容を学んでおいてもらえれば、より理解度が高まる可能性がある。同時に、授業内で扱う内容にも余裕ができるので、より丁寧に分析手法やソフト操作を説明し、期末レポートに向けた分析においてもデータ収集の段階からより本格的な分析体験をしてもらうことができるかもしれない。

あるいは、需要がどれほどあるかは分からないが、より高度にデータ分析を学びたい学生に対して、より数学的な解説を行う授業や、より高度な手法や因果推論の方法を学ぶ授業、RやPythonでの分析を試みる授業などへの展開も考えられる¹⁷。

また、データ分析の知識を前提として、テキスト分析などの手法を教える授業にも展開できる。テキスト

分析は分野横断的で汎用性の高い手法であり、身近なところでは口コミや自由回答の分析などで活用されている。また、テキスト分析では機械学習を用いた手法も盛んに活用されているため、本授業では欠けていた「AI」の部分について扱うことができるかもしれない。

以上のような複数科目の存在を想定した展開可能性は様々に考えられるが、いずれも現存しない授業であり、現状のリソースではこのような展開は困難であろう。したがって、もし本授業の試みを継続するならば、異なる方向性での工夫が求められる。

本節のはじめで述べたように、データ分析の基礎的な授業として、本授業には1科目完結の講義として一定の教育効果があったと思われる。また、リテラシーレベルのモデルカリキュラムおよび認定制度との関係においても、「基礎」「導入」「心得」などの項目には対応できていると考えられる。

以上を踏まえると、現状で対応可能な本授業の大きな課題は、受講者数に制限をかけざるを得ない現状かもしれない。授業の設計上やむを得ない面があるとはいえ、講義科目としてはあまりに受講可能な人数が少ない。

実習を伴い、ソフト操作のフォローや個別フィードバックが必要な授業であるため、既に述べたように本授業は2020年度、2021年度ともに受講者数に制限をかけてきた。受講制限は潜在的な受講希望者の履修登録を抑制してしまった可能性もある中、各年で予定した定員を超過する履修登録があり、10名程度の履修希望者には受講してもらうことができなかった。当然ながら、より多くの学生にデータ分析に触れてもらうためには履修制限はかけないことが望ましい。

本授業のような分析の実習を伴う授業では、受講生のフォローをTAやSAと協力して行うケースがある。そこで、安易な発想ではあるが本授業でも授業実施の補助をTAやSAに依頼する案はある。しかし、TA候補となる大学院生は必ずしもデータ分析の手法

を用いるわけではないため、TAの活用は難しい。一方、SAに関しては、例えば本授業を「秀」で単位取得した受講生について、次年度のSAを依頼するようなリクルート方法は考えられるかもしれない。

あるいは、データ分析の原理と活用事例を扱う講義科目と、実習と各自の分析に専念する演習科目に切り分けることも考えられる。この方法では講義科目で多くの受講生を受け入れることができるが、結局のところ演習型授業の方で履修制限をかけざるを得ないことになる。

本授業の来年度以降の開講は未定であるが、今後の展開としては、現行の内容と質をブラッシュアップしながら、多くの学生が受講可能になる工夫を検討していきたい。SAの活用は1つの解決策であるが、具体的にどのようにリクルートとするのか、SAの担当可能業務内でどのような補助をしてもらうのか、などをよく検討する必要がある。

最後に、困難なデータ分析に挑戦して見事に目標をはるかに上回るパフォーマンスを見せ、有益なフィードバックをしてくれた受講生達に心より感謝したい。

参考文献（五十音順）

- 浅野正彦・中村公亮（2018）『はじめてのRStudio: エラーメッセージなんかこわくない』オーム社。
- 浅野正彦・矢内勇生（2013）『Stataによる計量政治学』オーム社。
- 浅野正彦・矢内勇生（2018）『Rによる計量政治学』オーム社。
- 北川源四郎・竹村彰通 編（2021）『教養としてのデータサイエンス』講談社。
- 小宮あすか・布井雅人（2018）『Excelで今すぐはじめる心理統計—簡単ツールHADで基本を身につける—』講談社。
- 数理・データサイエンス教育強化拠点コンソーシアム（2020）『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～データ思考の涵養～』
< <http://www.mi.u-tokyo.ac.jp/consortium/model-literacy.html> >（2021年9月3日最終アクセス）

¹⁷ ただし、内容を高度化・専門化すると、頻繁に利用される統計ソフトや分析手法が分野により異なるので、どの方面に内容を拡張するかという問題は生じる。

数理・データサイエンス教育強化拠点コンソーシアム

(2021)『数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラム～AI×データ活用の実践～』<http://www.mi.u-tokyo.ac.jp/consortium/model_ouyoukiso.html>（2021年9月3日最終アクセス）

統合イノベーション戦略推進会議（2019）『AI戦略2019～人・産業・地域・政府全てにAI～』<<https://www8.cao.go.jp/cstp/ai/index.html>>（2021年9月3日最終アクセス）

文部科学省高等教育局（2021）『数理・データサイエンス・AI教育プログラム認定制度実施要綱細目』<https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm>（2021年9月3日最終アクセス）