

# 「探究」におけるデータサイエンスの活用可能性の模索

—社会調査の視点から—

滋賀大学 データサイエンス学部 伊達平和

「探究」は戦後第9回目の改訂となる学習指導要領において「総合的な学習の時間」を継承する形で導入された。データサイエンスはこの「探究」との相性はいいものの、そのまま現場に導入するには課題も多い。データサイエンスを応用する一分野に社会調査があるが、「イメージがしやすいこと」「方法論がある程度標準化されていること」「データサイエンスとの相性がよいこと」という点において利点があり、高校生の「探究」が求めている「自律性」に資する方法論として活用できる。

<キーワード> データサイエンス 探究 社会調査 アンケート

## I 授業科目としての「探究」の目的

本稿は高等学校における「総合的な探究の時間」の授業（以下「探究」）において、データサイエンスの活用の可能性について社会調査の視点から現場の教員に提言することを目的とする。本論に先だち、本稿の前提となる授業科目としての「探究」の特性について、今一度整理しておきたい。

授業科目としての「探究」は、戦後7回目の改訂となる学習指導要領（1999（平成11）年に告示、2003（平成15）年度から施行）においてはじめて登場した「総合的な学習の時間」の後継である。この学習指導要領のポイントはいわゆる「生きる力」の育成であり、その目玉でもあった「総合的な学習の時間」のねらいは学習指導要領によると以下の3つであった（文部科学省 1991）。（1）自ら課題を見付け、自ら学び、自ら考え、主体的に判断し、よりよく問題を解決する資質や能力を育てること。（2）学び方やものの考え方を身に付け、問題の解決や探究活動に主体的、創造的に取り組む態度を育て、自己の在り方生き方を考えることができるようにすること。（3）各教科・科目及び特別活動で身に付けた知識や技能等を相互に関連付け、学習や生活において生かし、それらが総合的に働くようにすること。つまり一言で表すと、生徒自ら主体的に問題を解決する態度・能力を育み、教科横断的に知識や技能を横断的にかつ総合的に働くようにするということが求められていたのである。

「探究」はこの「総合的な学習の時間」の延長線上にあり、戦後第9回目の改訂となる学習指導要領（2018（平成30）年に告示、2022（令和4）年から施行）によって「総合的な探究の時間」に名称が変更されたものである。「総合的な学習の時間」からの改訂内容や目標とするところは文部科学省の『高校学校学習指導要領（平成30年告示）解説 総合的な探究の時間編』に詳しい（文部科学省 2018）。要約すると、この「探究」の特質は「高度化」と「自律性」である。高度化とは「①探究において目的と解決の方法に矛盾がない（整合性）、②探究において適切に資質・能力を活用している（効果性）、③焦点化し深く掘り下げて探究している（鋭角性）、④幅広い可能性を視野に入れながら探究している（広角性）」を指し、自律性は「①自分にとって関わりが深い課題になる（自己課題）、②探究の過程を見通しつつ、自分の力で進められる（運用）、③得られた知見を生かして社会に参画しようとする（社会参画）」を指している（文部科学省 2018:9）。さらに教

科教育との差異において「探究」が目指すところは、特定の教科に留まらず横断的・総合的に「実社会や実生活における複雑な文脈の中に存在する事象を対象としている」（文部科学省 2018:10）とされている。つまり、「探究」は、「高校生自らの生活世界」をとらえなおすという「目的」のもと、各教科で学んだことを「手段」として働かせるということが目標とされている。

このようにあらためて「探究」の目標を概観すると、高校生や指導する教員が達成するのは難しいことでもあり、一筋縄ではいかない。一方で、大学教員からすると、以上に掲げられた目標は大学の日常の一部でもある。そこで社会調査の分野でデータサイエンスに携わっている一大学教員の視点から、「探究」におけるデータサイエンスならびに社会調査の可能性について整理してみたい。

## II データサイエンスと社会調査

### 1 データサイエンスとは何か

データサイエンスに対する社会的なニーズは日ごとに高まっている。職業としてのデータサイエンティストの重要性が指摘されてきたのは2010年ごろであったが（竹村・姫野・高田編 2019）、2023年の現在でもその流れは続いている。このようなデータサイエンスの社会的なニーズに呼応するように、データサイエンスに関する学部・学科の増加も著しい。データサイエンス学部を擁する大学は、2022年度までは滋賀大学、横浜市立大学、名古屋市立大学、立正大学、武蔵野大学の5大学であったが、2023年度からは、一橋大学ソーシャル・データサイエンス学部を筆頭に、データサイエンス関連の学部学科は17も誕生するという（朝日新聞 2020年11月26日）。このような社会状況において、もはやデータサイエンスや人工知能（AI）はバズワード化しているともいえる。その内実を知らずして、「必要と言われているから必要」というような風潮もあると聞くこともある。しかし、データサイエンスの内実を整理すると、近年突如として現れた謎の学問ではなく、これまでの学問の再編成という側面もある。

それでは改めてデータサイエンスとは何か。端的にいうと、情報学と統計学をもとにした新たな価値の創造である（竹村・姫野・高田編 2019）。まず情報学についてであるが、近年ではデータはビッグデータと呼ばれるように大規模化しており、効率的かつ適切に処理するためには情報学におけるプログラミングの知識や統計ソフトの操作への熟練が必要である。高校ではExcelが用いられることが多いが、多くのデータサイエンスの現場ではRやPythonが用いられている。そして統計学は、データを処理するための数学的技術である。例えば高校数学で学ぶ平均や分散、標準偏差、確率分布などの知識はこの統計学の初歩にあたる。ある変数の平均値を集団別に計算するというようなことを日常的にも行っているが（例えばクラス別平均値など）、これも統計学の実践的な活用である。データサイエンスの現場では、非常に大きなデータに基づき、回帰分析、因子分析、構造方程式モデリングなどより高度な数理モデルが使用されるが、形式的には同じようなことだと捉えることができる。

ただし、統計学と情報学の活用だけではデータサイエンスとは言えない。データサイエンスが強調するのは「価値の創造」であり、いわば「問題の解決策」をデータに基づいて導き出すことに主眼がある。この「問題を解決する」志向性をもつという点において、データサイエンスと「探究」は強く接点を持っている。

データサイエンスは多くの学問分野の方法論的基盤でもある。自然科学においては実験的手法が用いられることは頻繁にあるし、社会科学においても、たとえば心理学実験のみならず、人間や社会的事象の観察データの分析にデータサイエンスが用いられている。人文科学はデータサイエンスと接点が少ないと思うかもしれないが、テキストの処理はデータサイエンスの得意とするところであり、文学に対してデータサイエンスを応用することもある<sup>1)</sup>。

筆者の専門は社会学であるが、人間の意識や行動の分析には、データサイエンスが用いられている。この社会学が解こうとしている課題は決して日常生活から遠いものではない。例えば「幸福度が高いのはどのような人か」「どのような男性が家事・育児参加に積極的か」「どのような家庭背景がある人が大学に進学しやすいのか」「結婚はどのような人がしやすいのか」など、普段私たちが日常会話でしているような話題は、ほとんどがデータサイエンスによってアプローチ可能である。実際に先述したような問いに関するデータサイエンスを用いた研究の蓄積は非常に厚い。

データサイエンスの日常生活のかかわりについてはいくら強調してもし過ぎることはない。起床時にスマートフォンを開くときの画像認証や指紋認証の技術、天気予報を確認する、ベッドから起きて歯を磨くなら歯磨き粉の有効成分の根拠、朝食の牛乳とシリアルが家庭に届くまでの最適な経路の選択や商品の品質管理などもデータサイエンスが関わっている。起床してから家を出るまでにどのようなデータサイエンスが関わっているかを指摘するだけでも、この紙幅を埋め尽くしてしまうかもしれない。つまりデータサイエンスは、普段私たちの目には見えていないだけであり、「探究」が目標として定める「自分にとってかかわりが深い課題」を解くための方法論としては、適切なアプローチであるといえよう。

## 2 データサイエンスを「探究」に応用するときの課題

一方で、データサイエンスを高校の「探究」に応用するには多くの課題があると考えている。ここでは、これまで筆者が高校の「探究」のゲスト講師の経験を通した「肌感覚の報告」に過ぎないが、指導する側の教員の課題、指導される側の生徒の課題に分けて整理したい。なお『「探究」における課題の把握』という課題それ自体もデータサイエンス的なアプローチで整理することのできる問題であるが、これについては今後の課題としたい。

まず教員の課題<sup>2)</sup>について、データサイエンスを活用した「探究」の授業には、先述したように、データサイエンスの基本的な2つのスキルである統計学の知識と情報学の知識が必要だが、多くの先生がこれらを学んだ経験がないか、あるいは学んでいても実践的に使いこなすことができないという現状があると考えられる。Excelを利用して平均・分散のような基礎的な統計量を算出する、クロス集計を行う、散布図を出力し回帰直線を引くなどができるだけでも、データサイエンスを「探究」に導入することができるのだが、高校では教員の科目の専門性が高いため、教員一人一人がデータサイエンスの知識を持って生徒の指導にあたるのは難しいだろう。この点、数学あるいは情報を担当している教員が適しているものの、負荷が集中することになる。すなわち、教員とデータサイエンスを考えた時に課題になるのは、教育学部の教職課程に、あるいは教員研修の中に、いかにデータサイエンスを盛り込んでいくか、ということになる。

次に生徒の課題であるが、先述したように、数学では高校生は統計学の基礎を学んでいる。しかし「探究」の授業での活用ができるかと言われると、必ずしもそうではない。単純に忘れていたということも往々にしてあるのだが、数学の問題を解くことと、実際のデータの分析の間には大きなギャップがある。同じことは統計学を学んだ大学生が実データを取り扱う際にも起こるので、高校生にとってはなおさらともいえる。また高校生の場合は、そもそも数学を入試で使う予定がなければ、学んだことを覚えてはいないだろう。

このように、データサイエンスを「探究」に導入することを考えた時、様々な課題がある。そのような中で、データサイエンスの活用はいかに可能であるのか。次節では筆者の専門である「社会調査」に引き付けて考えてみたい。

### 3 社会調査の利点

ここではデータサイエンスの「探究」への応用可能性について、社会調査の観点から考える。まず社会調査とは社会を認識するための様々な方法論のことであり、質問票に基づく量的な調査、そしてインタビューやフィールドワークに基づく質的調査の二種類がある（伊達・高田 2020）。本稿ではいわゆる「データサイエンス」と親和性の強い量的調査に絞って記述するが、本来ならばインタビューやフィールドワークなどの質的調査も「探究」に応用できる（さらに言えば「質的調査」のデータにもデータサイエンスを応用できる可能性もある）ことは指摘しておきたい。

質問票に基づく量的調査の「探究」における利点については以下の3点であると考えている。それは「イメージがしやすいこと」「方法論がある程度標準化されていること」「データサイエンスとの相性がよいこと」である。

まず「イメージがしやすいこと」についてである。質問票に基づく量的調査は一般的には「アンケート」として認識されている。高校生や教員は、アンケートに回答した経験はそれなりにあるだろうし、高校生であれば中学生の時に作った経験、教員であれば日々の業務の中でアンケートそのものを作成したことがあるかもしれない。経験があるだけでやることのイメージがつきやすく、その意味で導入しやすい方法論である。これは良し悪しでもあるのだが「とりあえずやってみる」という感覚でも取り組むことができる。イメージが付きやすい分、「探究」が目標としている自律性「自分にとって関わりが深い課題になる（自己課題）」を満たすことにも寄与する可能性が高い。

次に「方法論がある程度標準化されていること」である。アンケート調査は一般向き（あるいは高校生向き）に解説書が多く出ている。解説書にはアンケート調査の手順が丁寧に書かれているが、筆者の場合、問題意識の設定→調査テーマの設定→リサーチクエスチョンの設定→仮説の設定→調査項目の設定…以降、調査・分析・報告と教えている。概ねどのような解説書でも類似の手順で紹介されているだろう。もちろん、それぞれのフェーズにおいて気を付けるポイントが少なからずあるが、手順が標準化していることは、一度調査の手順さえつかんでしまえば、毎年同じように指導を行うことができるという利点がある。また実践結果が蓄積されていけば、生徒も先輩の事例を参考にして、同じ手順でできるというのもよい点だろう。

最後に「データサイエンスとの相性が良いこと」である。アンケートでは様々な概念を問と質問のセットに変換して数値化する。例えば意識であればある項目への賛否を1～5の5段階で尋ねる、行動であれば具体的な頻度や時間を尋ねることで、平均や分散、相関係数といった統計量を算出することができる。例えば男女差を見たいのであれば、性別の平均値を比較したりクロス集計を行う。ある行動の背後にある意識の影響を見たいのであれば、相関係数を算出したり回帰分析を行う、といった具合である。もちろん、データ処理の方法はどこまでも高度化することができるが、初歩的な分析方法は限られているため、ある程度の目標を達成することができるだろう。仮説によっては単純な集計をするだけでも、議論を進めることは可能である。

## Ⅲ 「探究」における社会調査の活用可能性

### 1 大学における「社会調査」の実践例

筆者は大学では「社会調査」を教えている。ここでは、大学での取組み事例を紹介することで、「探究」における社会調査の活用可能性について考えてみたい。

量的調査の肝は調査票の作成である。この調査票の作成は先述したように、問題意識の設定→調査テーマの設定→リサーチクエスチョンの設定→仮説の設定の手順で進めていく。それぞれ内容とポイントについて

表1に整理する。

表1. 調査票を作成する上で前提となる4つの要素

	内容	ポイント
①問題意識	何を問題だと考えているのか	・問題意識は「身近なこと」からスタートするのが良い。 ・現状に対する認識、理想的な状況の把握、その2つのギャップから問題意識は生まれることを理解する。
②調査テーマ	問題意識に関する様々な現象のうち、何を調査のテーマ(対象)とするか	・問題意識に関連する現象は複数あるので絞る必要がある。 ・絞っていくときの基準として①できること、②やりたいこと、③もとめられていることの3点を満たすものにするとうまくいきやすい。
③リサーチクエスチョン	調査を通じて明らかにしたい問い	・調査を通じて明らかにできる問いになっていることが大切。 ・「べき論」(〇〇するべきか)に基づく問いを立てることがあるが、「するべきかどうか」は調査からは直接明らかにできないので注意。
④仮説	問いに対する仮の答え	・仮説を立てることができれば、その要素をそれぞれ質問と選択肢に落とし込むことで調査票が完成する。 ・仮説はなくても調査は可能だが、仮説を設定していた方が調査を効率的に行うことができる。

これらの手順のポイントについては表中に記載しているので省略するが、全体として重要なのは、それぞれの要素が、論理的に入れ子構造になっていることである。実際に自分でやったり、生徒が提示したのを見たりするとわかるが、①から④までの論理がばらばらになっていることがよくある。問題意識とリサーチクエスチョンがつながっていなかったり、リサーチクエスチョンに対する仮説になっていなかったり、さまざまであるが、この4つの要素をきちんと論理的に言葉で表現するためには国語力が必要である。その意味で「探究」における教科としての国語の重要性は極めて高い。

またもう1点大切なポイントとして強調しておきたいのは、高校生にとって身近なことを取り上げることである。例えば、「問題意識をあげよ」となると、「少子化を解決する必要がある」「地球温暖化が問題だ」といったような、いわゆる社会問題をあげることがあるだろう。そのような大きな問題は表中で示したように①できること、②やりたいこと、③もとめられていること、の3点を満たさない可能性が高い。もちろん大学の研究レベルでは社会調査を用いて「社会問題」を解決することが志向されるが、高校生の「探究」は教科の目標にも掲げられている通り、身近なレベルの問題にアプローチできることを重視するほうが良いのではないかと考えている。言い換えると、高校生の身近で自由な発想をいかに調査に仕立てることができるかが、指導のポイントでもある。

表2に大学生が「問題意識、調査テーマ、リサーチクエスチョン、仮説についてあなたの問題意識に基づいて書きなさい」という課題に取り組んだ時の一例を紹介する。①はコロナ禍における stay home の意識、②は地域環境の美化活動と公共的な性格が強い。しかし③はダイエット、④は「小さいバックを持つ女はいい女なのか?」という考え方についてであり、彼/彼女らの身近な問題意識からテーマを設定している。特に④はかなり個人的な問題意識であるが、仮にこれがマーケティングリサーチのような文脈で提示されると考えると(もちろん仮説はもっと丁寧に練る必要があるが)、ビジネスの現場でも調査されるような内容であるともいえる。

このように問題意識から仮説まで構築できると、調査票の作成はそれほど難しくはない。②の例で言えば、まずは自治会清掃に参加する頻度について尋ね、次に行動の原因となるような問を設定する。仮説にならうと世帯収入について、性別について、余暇時間について尋ねる。分析の仕方はいろいろ考えられるが、例えば男女別に参加頻度の平均値を比較する、世帯収入をいくつかのカテゴリに分割して参加頻度の平均値

を比較するなどの方法があるし、本稿では詳しく扱わないが、より高度なデータサイエンスの手法を用いるのであれば、重回帰分析によって関連性の大小について比較するようなこともできるだろう。

表2. 学生が考えたテーマの例

番号	問題意識	調査テーマ	リサーチクエスチョン	仮説
①	コロナの感染拡大防止でstay homeが求められている	stay homeに対する意識	どのような人々がstay homeを続けられるのか	インドア系の趣味を持っている人はstay homeを続ける意欲が高い 都会に住んでいる人はstay homeを続ける意欲が高い
②	地域の環境美化活動に参加する人数を増やして、活動を活発にすることで、きれいな町にしたい。	清掃に参加する人の特徴について	どのような人が自治会清掃に参加するのか	世帯収入の多い人ほど自治会清掃に参加する 労働時間が短い人ほど自治会清掃に参加する
③	ダイエットに成功したい	ダイエットに成功する人と成功しない人の違い	どのような人がダイエットに成功するのか	運動経験のある人はダイエットに成功しやすい 誰かと一緒に住んでいる人はダイエットに成功しやすい
④	小さいバッグを持つ女はいい女という風潮が気になる	小さいバッグを持つ女性の特徴	どのような人が小さいバッグを持っているのか	整理整頓が得意な人ほど小さいバッグを持っている こだわりが強い人ほど小さいバッグを持っている 心配性な人ほど小さいバッグを持っていない

注) 学生が考えた内容について、表現を一部修正して掲載

この大学生の事例で強調しておきたいのは、「身近」なところから問題意識を設定し、その問題にアプローチするという点で社会調査そしてデータサイエンスは威力を発揮するという点である。「探究」の改訂の意図を今一度確認すると、「高度化」と「自律性」であった。特に「自律性」という観点で考えると、社会調査は「自分にとってかかわりが深い課題になる」「自分の力で進められる」という点は言うまでもなく、さらに「得られた知見を生かして社会に参画しようとする」という点も「身近なテーマで進められる」ことから達成しやすいのではないだろうか。

## 2 「探究」における社会調査の限界

これまで、社会調査を軸に「探究」におけるデータサイエンスの活用可能性について論じてきたが、もちろん限界もある。本稿を締めくくるにあたり、「探究」の授業における想定される限界について示しておきたい。

まずは仮説の設定に関することである。仮説が検証できることはアンケート調査の利点であるが、調査の対象や目的によっては仮説が設定できないようなこともある。例えば、調査対象の全体性を知りたい場合、ある現象のプロセスや過程を詳細に知りたい場合、調査対象者の主観的な意味づけを知りたいような場合である。そのような調査テーマやリサーチクエスチョンの場合は質問票に基づく量的な調査よりも、インタビューやフィールドワークといった質的な調査の方が適切である場合がある。

次は調査テーマに関することである。社会調査は社会とそこで生きている人間の意識や行動に対して応用される方法論である。そのため例えば自然科学的な問題関心（物理学や生物学のような領域）、また人文科学的な問題関心（歴史や言語に関する領域）には対応できないことが想定される。そのような問題関心で探究を行う場合は、その領域の専門家が必要であろう。この点に関しては教科教育の先生の力を借りる、あるいは高大連携といったやり方で専門家や大学生・大学院生との協働が必要であろう。

最後にデータサイエンスのレベルについてである。今回データサイエンスの活用という点で具体的に言及したのは、高校数学でも応用可能な統計学の初歩的な分析方法であった。高校教員がそれぞれの専門性を持ち、データサイエンスで用いられる情報学と統計学の訓練を受けているわけではないという現状がある中で、本稿が言及するのはデータサイエンスのいわゆる初学者に限定した提言ともいえる。しかし、データサイエンスの高度な分析手法はほかにも沢山あり、中には人工知能（AI）の活用も含まれる。もし高校の内部

に指導できる教員がいたり、近隣の大学から指導者を招くなど、技術的な部分を補完できるのであれば、さらなる高みを目指して、データサイエンスを活用してほしい<sup>3)</sup>。

#### 注

- 1) 例えば樋口耕一氏の開発したテキスト分析ソフトウェアのKH Coderでは、夏目漱石の『こころ』の分析例がチュートリアルとして解説されている(樋口 2020)。
- 2) 教員の課題について指摘している研究報告は複数ある。例えば本田(2016)は研究型アクティブラーニングの現状と課題について、教育委員会・校長・教員・教員免許取得志望学生に対する調査を実施している。調査結果によると、校長の認識として、研究型アクティブラーニングの実践的な指導力を持つ教員の不足、指導に必要な予算・設備・時間の不足があげられ、また教員の認識としても実験の設計やアンケートやインタビューの社会調査の手法の能力不足があることが指摘されている。また「探究」に注目して調査を行ったものに蒲生(2018, 2019)がある。例えば蒲生(2018)は学校間での取り組みの格差について指摘しており、この点についても課題であろう。
- 3) ただし科目としての「情報Ⅱ」においては、教員研修用の資料として、回帰分析、主成分分析、分類や機械学習といった高度なデータサイエンスの手法が紹介されている(文部科学省 2020)。数学だけでなく情報とどのように領域横断的な学びを深めていくか、今後の実践報告が待たれる。

#### 参考文献

- 朝日新聞DIGITAL「データサイエンス系学部学科が続々誕生 就職好調 数学は必須？」2022年11月26日 (<https://www.asahi.com/articles/ASQCT66HVQBOLZU00H.html>) .
- 伊達平和・高田聖治, 2020, 『社会調査法』学術図書出版社.
- 蒲生諒太, 2018, 「全国高等学校「探究的な学習」に関するアンケート調査—探究先進校と一般校の比較検討」『同志社女子大学教職課程年報』1, pp. 44-62.
- , 2019, 「学校現場が直面する「探究的な学習」の課題について—全国高等学校「探究的な学習」に関するアンケート調査(2)」『同志社女子大学教職課程年報』2, pp. 3-12.
- 樋口紘一, 2020, 『社会調査のための計量テキスト分析【第2版】』ナカニシヤ出版.
- 本田由紀, 2016, 「研究型アクティブラーニングの現状・課題・可能性」『東京大学大学院教育学研究科紀要』56, pp. 245-262.
- 文部科学省, 1991, 『高等学校学習指導要領』(<https://erid.nier.go.jp/files/COFS/h15h/index.htm>) .
- 文部科学省, 2018, 『高等学校学習指導要領(平成30年告示)解説 総合的な探究の時間編』.
- 文部科学省, 2020, 『高等学校情報化「情報Ⅱ」教員研修用教材』.
- 竹村彰通・姫野哲人・高田聖治編, 2019, 『データサイエンス入門』学術図書出版社.