

情報 I における問題解決の手段としての「データサイエンス」

京都産業大学附属中学校・高等学校 教諭
森本 岳

1. 情報 I の各領域と問題解決の関係性

情報 I がよいよ始まる。ご存じの通り、情報 I の目玉は「情報デザイン」「プログラミング」「データサイエンス」の3つで、指導要領においても、「第2章 コミュニケーションと情報デザイン」「第3章 コンピュータとプログラミング」「第4章 情報通信ネットワークとデータの活用」となっている。3領域の具体的な取り扱いに関しては、指導要領に「第1章 情報社会の問題解決は、第2～4章との関連に考慮するものとする。生徒が情報社会の問題を主体的に発見し明確化し、解決策を考える活動を取り入れるものとする。」「情報技術を活用して問題の発見・解決を行う学習活動を通して、問題の発見・解決に向けて情報と情報技術を適切かつ効果的、創造的に活用し、情報社会に主体的に参画し、その発展に寄与するための資質・能力を育成することを目指す。」と明記されている。つまり「情報デザイン」「プログラミング」「データサイエンス」は、それぞれを個別に扱うのではなく、“問題解決”という探究の中で解決手段としてこの3領域を活かすことのできる力をつけることが重要で、3領域で学んだ知識・スキルが、目の前のリアルな事象に対して適切かつ自在に応用できるという状態を目指している。

実際、「情報デザイン」はUIやUXなどが抱える問題を解決する手段として使える他、問題解決の過程で他者に的確に情報を伝える（プレゼンや企画書・ポスター作成など）にあたって重要になる。

「プログラミング」は問題の生じる過程をモデル化することによる、より深い分析や、プロトタイプを様々な条件下でシミュレーションをすることで、より最適な解決方法を導くことができる。サステナブルなイノベーションには重要な要素である。

最後に「データサイエンス」は統計データやアンケートデータなどを活用し問題を細部まで分析したり、相関関係や因果関係を明らかにすることを可能にする他、実験やシミュレーションによって得られたデータを分析することで研究をさらに深められる。

以上のことから、これら3領域をひとつひとつ切り離して個別の課題として取り扱うのではなく、ひとつの大きな問題解決プロジェクトの中で、これらの解決方法の長所短所を理解した上で、うまく組み合わせ活用する経験こそが重要であると考え、本校では3年前から、その取り組み方法について試行錯誤をしている。今回は、問題解決の手段としてのデータサイエンスという観点で実践を紹介する。

2. データサイエンスと問題解決

問題解決を含む探究では、まず「問題が起こるメカニズムの解明」を行い、そのうえで「考えた創造的な解決方法の検証」をする必要がある(図1)。加えて、どんな研究領域も、すでに他者によってある程度明らかにされている“既知の領域”と、誰も明確な答えをみつけない“未知の領域”とが存在する(図2)ため、探究においては後者を“科学していく”必要がある。円の外側を科学する手段には「統計データ」「独自アンケートで収集したデータ」「独自の実験

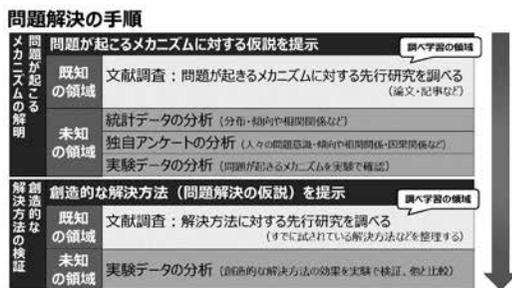


図1 問題解決の手順



図2 既知の領域と未知の領域

データ」の分析があるが、それぞれ明らかにできる側面が異なるため、実際の問題解決においては、未知の領域を明らかにできるよう、その特徴に注意しながら組み合わせを選ぶ必要がある。

次に「統計データ」「独自アンケートで収集したデータ」「独自の実験データ」の特徴をふまえ、どのように科学していくかを記すが、データサイエンスを活用した問題解決は、どの手段においても観察や実験に基づく実際のデータを元にパターンを見つけ出してルール化する“帰納法”の形式を取る。

3. 統計・アンケート・実験の違いと活用

「問題が起こるメカニズムの解明」「考えた創造的な解決方法の検証」ともに、最終的には因果関係を証明できる根拠を導くのがゴールだが、まずこの3つのメリット・デメリットを整理しておく(表1)。

「統計データ」では国勢調査や各省庁・調査機関が出しているデータを活用する。都道府県や時期などを軸に2つのデータの相関を明らかにする(図3)。近年はインターネットで簡単にデータを入手できるため、手軽に日本全国をくまなく分析することができる。しかし、統計データで明らかにできるのは「相関がどれくらい強さであるかどうか」までであり、両者の因果関係までは証明できない。そこで因果関係に迫ることができる「アンケート調査」の併用が必要になる。生徒が行うということもあり、広範囲に及ぶアンケートは不可能だが、統計データの存在しないニッチな領域についてのデータも集められる他、ターゲットの考え方と行動とを結びつけて調査できるため、因果関係を導き出すことができる。アンケートによって因果関係が突き止められたら、「演繹法」によって、統計データから全国的にもこの関

係が成り立つことを証明することができる。

このように「問題の起きるメカニズム」が明らかになれば、それを解決するためにどの方法が一番効果的か、考えた解決方法を「実験」によって検証することで、現時点での最適解を導き出せる。

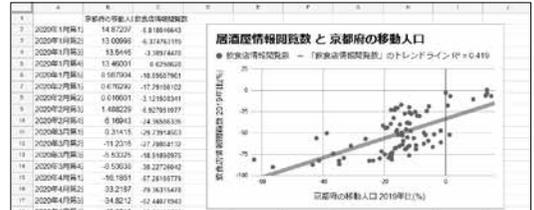


図3 京都における居酒屋情報閲覧数と移動人口
京都府における「移動人口」と「ネットで居酒屋を検索する件数」には正の相関がある。ただしこのデータだけでは、移動制限があったから検索数が減ったかどうかの決定打(因果関係の証明)にはならない。

4. 以上をふまえた本校での授業事例(高1生)

5月の授業でデータ分析(ヒストグラム・散布図・相関関係・因果関係・ピボットテーブルなど)の概要を学んだ後、課題形式で「ピボットテーブルを使ったアンケートのクロス集計」「ヒストグラムによる分析」「散布図による統計データの分析」などの練習を行い、6月にはスキル習得を確認するための実技試験を行う。7月までに、実際に自分が取り組む探究テーマの中で、学んだことをどう活かし研究を進めるか企画書にまとめ、夏休みに調査研究を行い、9月にプレゼンテーションを行う。このように体験的に知識スキルの定着が図れるようにしている。

参考文献

- 1) 文部科学省, 『高等学校学習指導要領(平成30年公示)』, 平成30年3月
- 2) 株式会社 Agoop, V-RESAS <https://v-resas.go.jp/> (2021年9月15日閲覧)

表1 統計データ・アンケート調査・実験データのメリットとデメリット

	メリット	デメリット
統計データ	<ul style="list-style-type: none"> 日本全国など広範囲で調査されたデータを活用できるため、<u>全国的な傾向や、地域ごとの傾向の違いを明らかにできる。</u> 適切にサンプリングされているデータを使うことができる。 	<ul style="list-style-type: none"> 都道府県ごとや年代ごとの相関関係をみつけることはできるが、<u>因果関係まではわからない。</u> 多くのデータが提供されているが、自分の欲しいポイントのデータで無いこともある。
アンケート調査	<ul style="list-style-type: none"> 自分の欲しいデータを集めることができる。 仮説に基づいて設問を作ることができ、<u>考え方と行動を結びつけることができるので、因果関係にまで言及できる。</u> 	<ul style="list-style-type: none"> 統計データほど広範囲で大規模な調査を行うのは現実的ではない。 サンプリングに偏りが出たり、設問が適切でない可能性がある。
実験データ	<ul style="list-style-type: none"> どのように問題が起こるのか、<u>どの解決方法が一番効果があるのか、を直接実験で検証することで、何よりも有力な根拠(一次資料)を得ることができる。</u> 	<ul style="list-style-type: none"> 生徒でも取り組める実験であることが大前提になるため、できることが限られてしまう。 適切に測定できているかなど実験の質が結果に影響を及ぼす。