

卒業報告書

石原みやび

miyabi.ishihara@gmail.com

2023年12月に統計学の博士課程を修了しました。大学院では、異常気象データの分析手法や、衛星画像データを数値データに変換する方法、そのデータを用いて気候リスクの経済的影響を調べる研究に取り組みました。大学院生活では、データを目の前に模索と試行錯誤を重ねる悶々とした日々が多くありましたが、研究内容を提出することができ、船井情報科学振興財団の継続的なご支援、共同研究者、家族、友人の皆さまに心より感謝申し上げます。

卒業後は統計学の教育に携わりたいと思い、現在はペンシルバニア大学の統計学・データサイエンス学科で講師を務めています。学部生、博士生、MBA生を対象に、統計学入門、プログラミング、機械学習の授業を担当しています。ティーチングアシスタントのときとは異なり、教員になってからは、学期末に学生にどのような力を身につけてほしいかを考え、それをもとに授業内容を逆算して構成するよう意識するようになりました。どのように説明すれば伝わるか、どんな練習問題が理解の助けになるか、学生の学びをどう評価するかを考えます。また、学生から寄せられるさまざまな質問や要望に対応するには、根気強さも求められます。一方で、授業の準備や学生との会話を通じて、私自身の理解も深まっていくのを感じました。



フィラデルフィアは、1776年にアメリカ独立宣言が行われた場所として知られています。街はコンパクトで交通機関も通っているので、車がなくても生活できます。気候は関東地方に似ており、夏は湿度が高く、先日住宅街の草地でホタルを見かけました。



ペンシルバニア大学のキャンパス内にある、創設者のベンジャミン・フランクリンの像。学生の間には、ビジネススクールや医学部の学生を中心に、TAなど学内での仕事や実績づくりに熱心な風土があるように感じます。

今回の報告書では、統計学に関わる中で感じた、日常の中でデータを読み・扱うときに意識したいと思ったことについて書きたいと思います。

データの背後にある権力構造と利害関係

成績、IQ、大学ランキング、GDP、幸福度指標、クレジット（信用）スコア…。データは、統一された方法で測定されることで比較やモニタリングがしやすくなる一方で、複雑な現象を単純化する行為であると捉えることができます。「表し方」は唯一無二のものではなく、誰が・何を・どのように数値化・測定するかによって大きく異なります。

たとえば私自身、教員として学生の成績をつけるためには、課題・試験問題を作成し、評価基準を決め、どの得点範囲がAかBかCかを決めます。このとき、私が妥当だと考える評価基準には、私がこれまで受けてきた授業や教育経験がどうしても関係しています。評価の決め方は数多くあり、教材の内容や重視する観点が異なれば、教員ごとに評価の仕方も変わってくるでしょう。つまり、山ほどある評価方法の中から、どれか一つを選ばなければいけないのですから、データを取るという行為（この場合は学生を評価すること）は、ある種のバイアスを生み出す行為とすら言えます。何を・どうデータ化するかを決めるのは、ある特定の人（あるいはアルゴリズム）であり、その判断にはその人の思考や価値観（学習データや指示された目標）が反映されています。

成績には、学生の学びを促し、学修状況を可視化して教員と学生の双方が学び方・教え方を調整できるという利点があります。一方で、成績が学生の過度なストレスの要因となることもあります。というのも、成績は授業内で完結するものではなく、しばしば授業の枠や時間を超えて影響を及ぼすからです。たとえば、奨学金の受給可否、大学院出願、スポーツチームへの所属継続、就職活動の評価などに成績が評価の対象になっています。成績は本来、学びを促進するための手段として使われ始めましたが、現在はそれを超えて、経営側の意思決定を効率化するための優先順位づけの手段としても使われている側面があります。

一教員として、学生に対しこうした成績の側面に理解を示しつつも、「成績にとらわれすぎず、自分の学びに集中してほしいし、私の役割はそれをサポートすること」と伝えています。また、評価方法については、あらかじめシラバスで明示し、学期を通じた学修状況によっては成績をある程度底上げできるような余地を残しています。評価方法に対する共通認識をつくり、学生に寄り沿った対応ができるようにすることで、評価の完全なブラックボックス化・トップダウン化を避けるように意識しています。教員という立場になって、人を評価・測定することの背景や難しさを知る中で、私自身、他者が作った評価の世界観に過度にのめり込み、一喜一憂しすぎず、自立した姿勢でいたいと感じています。

本当とはなに

統計を含む科学的な方法論では、ある仮説を直接的に証明するのではなく、それを反証しようとする方法をとります。本来は、ありとあらゆる反証の試みに耐え抜いた仮説が、現時点で最も妥当な説明と見なされます。ただ、それも永久に正しいと保証されるわけではなく、新たな反証の方法によって将来的に否定される可能性を常にはらんでいます。

イメージとして、自動車の安全性を確認するテストが参考になるかもしれません。たとえば、目の前の自動車が本当に安全かを確認したいとき、ただ舗装された道をまっすぐゆっくり走らせるという試験を行っても、その車が安全である良い検証にはなりません。そんな簡単な試験では、安全でない車を炙り出すことはできないからです。むしろ、スピードを上げたり、でこぼこ道を走らせたり、重量を加えたり、長時間運転したりといった、ありとあらゆるストレス条件下にさらし（安全でないはずという反証行為をたくさん行い）、それでも問題が起きなかった場合に限って、安全性が高いと判断されます。ただ、その試験の範囲外で後に問題が発見されれば、その環境における安全性は反証されたこととなります。安全であるという説は、現時点で反証されていないという限定付きの状態にすぎず、新たな条件下では覆される可能性があります。

ある説が正しくないかもしれないという前提のもとで、それを反証できるような厳密な検証を積み重ねることによって、明らかに事実とは言えない説を一つずつ除外していきます。そうした行為の末に、ようやく、真実の輪郭がぼんやりと浮かび上がってくるような感覚です。

データ分析にも似たような側面があります。データの収集方法が数多くあるのと同様に、同じデータでも処理や解析方法が数多くあります。そのため、それぞれの研究には固有のバイアスがあります。やろうと思えば、意図的に自分の思想を支持する結果が出るように解析を逆算して行うことも可能ですし、このことについては実際に研究分野で長らく問題視されています。一つのデータ解析の結果だけで何かが完全にわかることは稀だと感じます。

たとえば、肺がんと喫煙の関係をめぐる研究では、かつて、女性は肺がんになりにくいという説がありました。しかし、アメリカの女性の喫煙率の上昇に伴い女性の肺がん発症率も増加したことから、少なくともその説は成り立たないことがわかります。イギリスのデータに基づく研究結果に対してイギリス人以外には当てはまらないという批判があっても、他の人種を対象とした研究によって反証されています。このように、世界中のさまざまなデータ、さまざまな解析の積み重ねにより、喫煙が肺がんリスクであるという説へと収束していきます。

一方で、どのような根拠が必要か、そしてどれほど強い根拠が意思決定に必要とされるかは、状況によって異なります。とくに公共政策のように人命や健康に関わる場面では、リスクが完全に立証されていないことを理由に何の対策もしないのは適切とは言えません。むしろ、データが少ない段階でも判断を下さなければならないこともあります。

本当のような嘘に溺れないための力は、生成AIの普及によってこれまで以上に重要になってきているように感じます。たとえば教育現場では、生成AIを用いて課題に取り組む学生が増えています。もちろん、生成AIは学びを支える強力なツールとなり得ますが、一方で、事実でない内容をあたかも本当のように提示することもあります。学びの初期段階にある人にとっては、こうした情報を鵜呑みにしてしまうリスクがあるように見受けられます。また、生成AIは人間に心地よく思わせるように説明するよう設計されているため、なおさら聞き手の警戒心が緩み、ぬるっと騙されやすくなってしまう可能性があります。実際、特定の企業や国に有利な情報の拡散、プロパガンダに近い返答が確認されつつある一方で、それを阻止する有効な手段が今のところない状況です。優しく語りかけるが巧妙な言葉や意識操作に対して、自分の中に生じたモヤモヤな気持ちに蓋をせず、一歩引いて確認をする姿勢が必要です。教育現場は、個人の感受性を尊重しつつ、結果を焦ることなく事実を確認するための思考の筋力をトレーニングする場として、貴重な環境だと感じています。

私自身、アルゴリズム技術の普及によってより顕著になりつつある社会の課題、事実確認の手法について、これからも考え続けていきたいと思います。これらのトピックにご興味があればぜひご連絡ください。