

## (3) 大学におけるデータサイエンス教育の在り方

### —本学の授業分析を通して—

教科横断的統計学

井上絵美, He Jiayi, Zhang Mingshi, 柴田未央, 松原心, Zhang xinyi

近年、情報化が進み、それに伴って「データサイエンス」の重要性が高まってきている。特に、教育分野では急速に動きが活発化しており、2021年度から文部科学省を中心に「数理・データサイエンス・AI教育プログラム認定制度 (MDASH)」が設立された。これまで制度自体を対象とした研究は多くあるものの、具体的な授業内容を対象とした研究は行われることはなかった。そこで本研究では、高等学校教員へのインタビューや本学で行われている授業を分析することを通して、大学におけるデータサイエンス教育の在り方について考察する。分析の結果、授業方法としては反転授業の現状と課題が明らかになり、そして授業内容としてもデータサイエンスの「活用」を促すために、演習問題の改善が必要であることが明らかになった。

Keywords : データサイエンス, 数理・データサイエンス・AI 教育プログラム認定制度, 統計, 反転授業

#### 1. 研究の背景と目的

近年、世界中で情報化が急速に進んでおり、日本国内でも2021年にデジタル庁が発足するなど、情報化に対応した動きが起こっている。その中で、注目されているのが「データサイエンス」である。お茶の水女子大学は「データサイエンス」について、「データサイエンスとは様々なデータを処理し、分析することで、そこから新たな知見を引き出すための手法であり、データサイエンスを駆使してビッグデータから新たな価値を創造する人材のことをデータサイエンティストと言います」と述べており、その「データサイエンティスト」の育成に向けて特に動きが進んでいるのが、教育分野である。例えば、高等学校では、2018年学習指導要領改訂により、数学科に「統計的な推測」(数学B)が追加されたり、情報科が必修化されるなどしている。また、2021年度から「数理・データサイエンス・AI教育プログラム認定制度 (MDASH)」(以下MDASHと表記する)が設けられ、数々の大学が認定を受けている。

本研究では、大学におけるデータサイエンス教育に着目するが、これまでの大学教育についてのデータサイエンス研究は、MDASHが最近設立された制度であることから、MDASHという制度自体を分析するもの(田口(2022)など)や、MDASH認定大学の教育プログラムの構成について分析するもの(新原(2023)など)が主であり、具体的な授業方法や内容については議論されることが無かった。これから、データサイエンス教育がさらに進んでいくことを考えると、

制度自体の研究やプログラムの構成についての研究だけではなく、具体的な授業方法や授業内容についての研究も必要になってくる。

そこで本研究では、MDASH認定校である岡山大学が教養教育科目として開講している「数理・データサイエンスの基礎」を分析対象として、授業の方法や内容について検討し、大学におけるデータサイエンス教育の在り方について考察する。

#### 2. 高等学校におけるデータサイエンス教育の現状

大学におけるデータサイエンス教育の在り方について考えるのに欠かせないのが大学の前段階である高等学校におけるデータサイエンス教育の現状についての把握である。

先にも述べたように、高等学校では2018年の学習指導要領改訂により、数学科の数学Bの内容に「統計的な推測」が追加されたり、情報科が必修化されるなどした。しかし、新学習指導要領は、2022年度入学の生徒から年次進行で実施されているため、新学習指導要領導入後の高等教育におけるデータサイエンス教育についての研究は少なく、先行研究からの分析は困難であると判断した。

そこで、本研究では、岡山県内の高等学校に勤める教員を対象にインタビュー調査を行うことで、高等学校におけるデータサイエンス教育の現状について明らかにすることとした。

##### 2-1 インタビューの対象と方法

岡山県内の高等学校で数学を教えている教員1名を対象にインタビュー調査を行った。インタビュー時

間は1時間程度で、データサイエンスと親和性の高い「統計的な推測」(数学B)の履修状況や実態、大学に求めるデータサイエンス教育などについて、構造化インタビューを実施した。

## 2-2 インタビューの結果

インタビューで尋ねた大まかな質問とその回答を以下に示す。

- ・「統計的な推測」は全員履修するか  
→普通科高等学校は特にそう。
- ・「統計的な推測」を共通テストで選択するか  
→恐らくその流れだと思う。大学によっては、二次に出すって言うけど、まあ、あれがメインで出るとはまずないから、何かと絡めて出てくる可能性はあるけど。まあ、要は共通テストで取れてことやね。
- ・「統計的な推測」を教えてみてどうだったか  
→ちょっとこっこの予想を超えたのが、夏休みの補習で終わるはずだったんよ。けど、終わらなんだ。(中略)内容が多いと思った。  
→模試とかで具体例が多いようなデータをやってみて、まあいろんなパターンを経験せんと、共通テストとかには太刀打ちできんかなあって気はする。
- 例えば、そのな、データサイエンスに力を入れる学校とかは、文部科学省とか県とかが、そういう何らかの授業を立てて、それに参加しなさいって言われたりしたら、普通のカリキュラムの上にそういうのを用意して、で、まあ、最近の流行りで言うと、教科横断的なことをしながら、やっていくんじゃないけど、まあ、それはほんとに、指定された一部の学校の話であって、それ以外の学校はそこまで力を入れるかっていわれたらそれはちょっとクエスチョンやな。

まず、「統計的な推測」は大学進学が進路の主な選択肢である普通科高等学校は履修し、基本的には共通テストも選択することが示唆された。また、「統計的な推測」の内容は量的には多いが、一方で共通テストを見据えた時には、教科書の内容のみでは不十分である可能性が示唆された。

さらに、高等学校の立場から、データサイエンスについて大学教育に求めるものは、高等学校の現状理解、そしてデータサイエンスがどのように生かせるのか、社会でどのように活用していくことができるのかといった点を考えることができるようにというものであった。これは、高等学校、特に普通科で

あれば、大学入試のための授業になる傾向になり、時間的制約も多いことから生まれる考えであることが予想され、その他の高等学校もある程度共通した現状なのではないだろうか。

- ・データサイエンスについて大学に求めるものはあるか  
→大学で習う、ゼミとかで研究することの基礎を高等学校ではやっていますよっていうことで、それを知ってもらいたいくらいかなあ。だけど、学校によってある程度格差があるから、そこはみながみんなそうじゃないんで。  
→高等学校の先生としては、まあ、将来的にやっぱなあ、社会に出て行くから、で、その距離は高等学校より近いじゃん、っていうことは、その、ほんとと将来社会に出て、役に立つ人間になるために、その、データサイエンスをどんな風に使うと、社会の役に立つのかとか、あの、みんながこう幸せに暮らせるんかとか、っていうようなことを大学の時にちょっとでも経験できたら、えんかなあって思う。

## 3. 現在の授業の分析

### 3-1 現在行われている授業の概要

教育学部の学生を対象として実施されている「数理・データサイエンスの基礎」は以下の流れで授業が進んでいく。

- 第1回:オリエンテーション, 教育とデータサイエンス
- 第2回:1変数の記述統計とその演習(反転授業)
- 第3回:標準化, 散布図, 共分散とその演習(反転授業)
- 第4回:2変数の記述統計とその演習(反転授業)
- 第5回:統計的因果推論の基礎

エビデンスに基づく教育実践・政策形成の理論

- 第6回:AIと機械学習, 機械学習の考え方と評価
- 第7回:EXCELによる統計解析とその演習

以上の流れで、「データを多面的に読み解く力を養う」こと、そして「データサイエンスの知識・スキルが活用される具体的場面について理解を深める」ことを目標として授業は進んでいく。

本研究では、以上の目標を念頭に置きながら、第2~4回の反転授業で行われる統計の学習についての講義を対象とし、分析を進めていく。

### 3-2 方法について

この授業では、反転授業が特徴であると言える。長

期的な反転授業の効果についての先行研究は多い。加藤 (2019) は、反転授業は従来方式の講義と比べて、予習の精度を高め、記憶の定着に貢献する可能性が示されたと指摘した。また、梅澤 (2018) は反転授業でのグループ活動の効果を検証し、従来型の対面授業と比較してグループ分け対面授業に効果があることを示した。このように、先行研究では、反転授業は学生の活発レベル、授業の参与レベルを高めることを示した。

本節では、本学の反転授業の形で行われる統計学習授業の現状について分析する。分析の対象となる第2~4回の授業では、学生は大まかに以下の流れで活動する。

1. 予習動画を視聴し、予習プリントを作成する
2. 個人で演習問題を解く
3. グループで演習問題を解く
4. 確認テストを解く

授業を分析するにあたって、授業中の学生の様子について確認するための調査用紙を作成し、授業をサポートする TA・SA に記入の協力をして頂いた。調査用紙の内容は、予習動画や演習問題についての意見、個人活動でのグループの演習問題の到達度、グループ活動でのグループの演習問題の到達度を TA・SA に評価してもらうものである。個人活動・グループ活動でのグループの演習問題の到達度について、現在の授業内で行う演習問題は以下の構成になっており、どのレベルまで解答できているのかを評価したのが到達度である。

レベル A : 定義の確認

レベル B : 定義通りの計算問題

レベル B+ : センター試験過去問など

レベル C : データを多面的に読み解く問題

以上を踏まえ、TA・SA の調査用紙に基づく以下のクロスセクションデータベース(表1)を作成した。

表1 各変数のまとめ

Variable	N	Mean	Media	sd	Max	Min
			n			
Score	70	40.33	42.05	6.80	51.00	20.25
Group	70	/	/	/	/	/
Room	70	0.62	1.00	/	1.00	0
individu al	70	2.38	2.00	0.86	4.00	1.00
group	70	3.00	3.00	0.80	4.00	1.00
member	70	3.93	4.00	0.73	5.00	1.00

注 : Variable は変数の名前、N はサンプル数、Mean は平均値、Median は中央値、

sd は平均誤差、Max は最大値、Min は最小値

ここで、Score は各グループの予習プリントの成績の平均値である。予習プリントとは学生が授業前に予習動画の内容を踏まえて自分なりにまとめたもので、TA・SA によるルーブリック評価で成績が付けられる。Group は各グループの名前(基盤1, 基盤2...) である。Room は教室ダミー変数である。本授業は2つの教室で実施されており A 教室を0, B 教室を1とする。A 教室は、小学校教育・幼児教育の学生、B 教室は中学校教育・特別支援教育・養護の学生が割り振られている。Individual は個人活動でのグループの演習問題の到達度、group はグループ活動でのグループの演習問題の到達度を示しており、レベル A まで到達したグループは1, レベル B は2レベル B+ は3, レベル C は4 いうようになっている。member はグループの出席者数である。

表1より、各グループの予習プリントの成績(Score)は、平均値と中央値はともに40前後だが、最大値と最小値の間に30の差があり、標準偏差も6.8であることから、グループごとに大きな差があることがわかる。これは専門性の違いからグループごとの知識の蓄積が異なることが要因として考えられる。「数理・データサイエンスの基礎」は教育学部の必修科目で、グループ分けを教育分野ごとに行っている。つまり、理系や技術系の学生にとっては高い成績を取りやすい科目となり、グループ間に大きな差が生まれる状況であると考えられる。

表2は、個人活動・グループ活動、それぞれの時間で各グループがどの程度到達したのかを表している。見てみると、個人の時間からグループの時間にかけて到達度1, 2のグループ数は減少し、到達度3, 4のグループ数は上昇している。つまり、グループ活動を通して、各グループの到達度が上昇して

表2 各到達度に達したグループ数

到達度	個人	グループ
1	1 1	3
2	2 7	1 2
3	2 5	3 3
4	7	2 2

いることが言える。

次に、関心のある変数 score, individual, group member を選択し、相関分析を行う(図1) score と individual, group との相関を見てみると、score と individual の相関係数は0.52、score と group の相関係数は0.53であり、どちらも正の相関が見られ

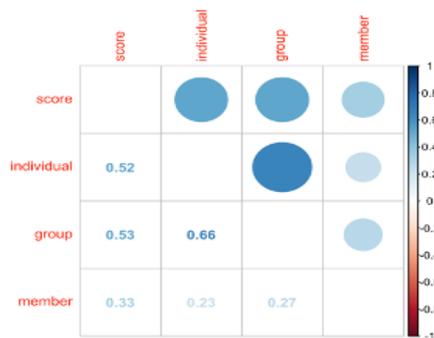


図1 相関図

る。つまり、現在の授業においては、予習プリントの成績が良くなると演習問題の到達度が高くなることが言える。

### 3-3 内容について

データサイエンス教育に求められるものについて確認すると、データサイエンスの定義の分析から、データをどのように「活用」するのが重要視されており、インタビュー結果から、高等学校教員の視点でも、大学のデータサイエンス教育に、「データサイエンスをどんな風に使うと、社会の役に立つのか」を考えることが大事と捉えられていることが分かった。そして、先にも確認したように、この授業の到達目標として、「データサイエンスの知識・スキルが活用される具体的場面について理解を深める」という目標がある。

そうした目標の達成のための問題として演習問題に位置付けられているのが、データを多面的に読み解く問題であるレベルCである。しかし、今年度の授業のSA・TAのアンケートより、レベルCに到達していない班が多く存在することが分かった。また、レベルCに到達していない班の多くはレベルB+までの問題の計算に躓いていることも分かった。この授業で扱う統計学の内容のほとんどが、高等学校数学Iにて学習済みの単元である。また、実際に統計を使う際、人間の手で計算することは少ない。そのため、この授業ではデータを多面的に読み解く問題に重きを置くことが良いと考えられる。また、授業終了後に行ったアンケート調査で学生に「あなたは今「データサイエンス」はどのようなものだと考えていますか？」という質問を行ったところ、「データの活用」や「具体的場面」という視点を持った回答は、全体の7.7%であった。データサイエンスの定義については、絶対正しいものはないが、大学の授業で目指す姿には到達できていない状況であると言える。この課題を乗り越えるためにも、演習問題の改善は必要になると考えられる。

他にも、現在の問題の中に、数問、問題の表現が適切でない問題が見られた。

### 3-3 結果まとめ

以上の分析から、授業方法については、反転授業という授業形態が、学生の学習に良い影響を与える可能性が示唆された。一方で、授業内容については、授業内で行う演習問題の内容が目標に対して適切でなく、改善の余地があることが分かった。

これらの分析結果を生かした授業を作成し、実践することができれば、これからの時代に求められるデータサイエンス教育に寄与できるのではないだろうか。

### 4. 今後の展望

本研究では、大学におけるデータサイエンス教育の在り方について、授業の方法と内容の2つの視点から考察し、現在行われている大学の授業の現状と課題、そして今後の大学におけるこれからのデータサイエンス教育の在り方を考察した。今後は、①高等学校の現状についての追加調査、②具体的な授業案の提案の2点に取り組む必要があると考える。本研究では、高等学校の現状を把握するために高等学校の教員1名にインタビューを行った。今回聞き取った内容が一定程度一般化できる内容であると仮定し、本研究を進めてきたが、各教員の意識の違いがあることも十分に予想でき、この点についてはまだ検討が必要であると考えている。また、本研究では大学の授業を分析することを通して、大学におけるデータサイエンス教育の在り方を示唆したが、それらを具体的にどのように活かしていくのかについては、今後検討が必要であろう。

### 参考文献

- 1) 田口敏行(2022)「数理・データサイエンス・AI教育の動向と在り方について—本学への応用を考える材料として—」『静岡産業大学情報学部研究紀要』第24号別冊
- 2) 新原俊樹(2023)「数理・データサイエンス・AI教育プログラムの実状—2021年度リテラシーレベル認定78校の事例から—」『日本教育工学会論文集』47(2), pp. 333-342
- 3) 梅澤克之, 石田崇, 中澤真, 平澤茂一(2018)「グループ分け反転授業の実授業への適用について」『経営情報学会 全国研究発表大会要旨集』PACIS2018, pp. 35-38
- 4) 加藤 研太郎, 高島 恵(2019)「基礎科目に対する反転授業の効果」『理学療法-臨床・研究・教育』26(1), pp. 29-35