

文系大学生のデータサイエンス教育における 理解度向上を目指した教授法の比較分析

河合 麗奈¹・伊藤 大河²・板橋 咲季³

¹正会員 共栄大学 国際経営学部 助教 (〒344-0051 埼玉県春日部市内牧4158)

E-mail:kawai@kyoei.ac.jp

²正会員 共栄大学 国際経営学部 准教授 (〒344-0051 埼玉県春日部市内牧4158)

E-mail:t-ito@kyoei.ac.jp

³非会員 デジタルハリウッド大学大学院 デジタルコンテンツ研究科 院生
(〒102-0062 東京都千代田区神田駿河台4-6 御茶ノ水ソラシティ アカデミア 3F/4F)

E-mail: g231tg2001@dhw.ac.jp

日本における「数理・データサイエンス・AI教育プログラム」では、リテラシーレベルが「大学等の全ての学生が身に付けておくべき素養」と位置付けられ、文系大学生でもデータサイエンスの学習が求められている。しかし、文系大学生は数学への苦手意識が強く、従来から実施されている数学の知識を前提として数式を活用する教授法では理解が困難であると推測される。

そこで本研究では、標準偏差をテーマに、数式、表、Excelを活用した3種類の模擬授業を文系大学生対象に実施し、どの教授法が適切なのかを調査した。その結果、表を用いた授業が最も評価を得られた。また、社会に出て役立つ知識が得られ将来に活かせると考えているのはExcelを用いた授業だった。そして、総合的な評価としては、Excelを活用した授業が最も文系大学生に効果的な教授法だと考えられる。

Key Words : data science education, humanities students, teaching method, comprehension

1. 緒言

情報社会が進展し、データが私たちの生活のあらゆる側面に浸透している。このデータ駆動の時代において、データサイエンスは、科学技術だけでなく、人文科学や社会科学の領域でもますます重要な役割を果たすようになっていく。データサイエンスとは、「アルゴリズムや統計などといった情報科学系の理論を活用してデータを分析し、有益な知見を見出すことを追究する新しいアプローチ」であり、統計学、情報理論、コンピュータサイエンスを組み合わせた学問である。そして、これらの知識を活用してデータから有用な情報を抽出し、新たな知識を創出することを目指している。この能力は、文系大学生にとっても非常に価値がある。なぜなら、将来的に直面すると考えられる課題や問題の多くは、データに基づいて理解し、解決することが可能だからである。

(1) 社会的背景

日本政府がまとめた「第5期科学技術基本計画²⁾」では、「超スマート社会サービスプラットフォームを活用し、

新しい価値やサービスを生み出す事業の創出や、新しい事業モデルを構築できる人材、データ解析やプログラミング等の基本的知識を持ちつつビッグデータやAI等の基盤技術を新しい課題の発見・解決に活用できる人材などの強化を図る」としている。

また、2019年に日本政府は、「人間中心のAI社会原則³⁾」と「AI戦略2019⁴⁾」を取りまとめた。人間中心のAI社会原則では、AI-Readyな社会(AIを有効かつ安全に利用できる社会を構築すること)への変革を推進する必要があることと、3つの理念、7つのAI社会原則が示されている。このAI社会原則の1つが「教育・リテラシーの原則」であり、全ての人が平等にAI、数理、データサイエンスの教育を受ける機会を持ち、AIの概要を理解し、正しく利用できる素養を身につけていることが望まれている。そのためにも、教育環境が個々の関心や力を活かす方向へ変化し、社会全体で教育環境の整備に取り組むことが重要であるとし、AI社会における教育とリテラシーの重要性を示している。AI戦略2019では、「我が国が、人口比ベースで、世界で最もAI時代に対応した人材の育成を行い、世界から人材を呼び込む国となること。さらに、

それを持続的に実現するための仕組みが構築されること」が第一の戦略目標と位置付けられた。そして、大学生を対象とした具体目標では、2025年度までに「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」することとしている。

(2) 教育的背景

文部科学省より数理およびデータサイエンスに係る教育強化の拠点校として選定された6大学（北海道大学、東京大学、滋賀大学、京都大学、大阪大学、九州大学）は、2017年度に「数理・データサイエンス・AI教育強化拠点コンソーシアム」を形成し、文理を問わず全国すべての高等教育機関の学生が、数理・データサイエンス・AIを習得できるような教育体制の構築・普及を目指して活動している⁵⁾。そして、データサイエンス教育に関するスキルセットおよび学修目標の公表や、数理・データサイエンス・AIモデルカリキュラムの公表などを実施している。また、内閣府、文部科学省および経済産業省は、2020年に「数理・データサイエンス・AI教育プログラム認定制度」を創設した⁶⁾。この制度は、デジタル時代の「読み・書き・そろばん」である数理・データサイエンス・AIに関する、大学（短期大学含む）・高等専門学校の正規の課程の教育プログラムのうち、一定の要件を満たした優れた教育プログラムを文部科学大臣が認定/選定することによって、大学等が数理・データサイエンス・AI教育に取り組むことを後押しする制度である⁷⁾。デジタル社会の基礎的な素養としての初級レベルの習得を目指す「リテラシーレベル」と、自らの専門分野において、応用・活用することができる応用基礎力を習得することを目指す「応用基礎レベル」の二段階に分かれており、特に優れたプログラムを「プラス選定」している。

リテラシーレベルは、大学・高等専門学校を卒業する全員（年間50万人想定）が習得する内容とされており、モデルカリキュラムでは、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）がコア学修項目（必須項目）、その他の数理・データサイエンス・AIを活用するための基礎的な知識・スキルが選択（オプション）項目となっている⁸⁾。つまり、文系大学生であっても、この数理・データサイエンス・AI教育プログラムのリテラシーレベルが身につけておくべき素養として位置付けられている。

(3) 文系大学生を対象としたデータサイエンス教育

文系大学生を対象としたデータサイエンス教育に関する研究も各種実施されている。例えば松尾ら⁹⁾は、文系大学生におけるデータサイエンス教育の在り方を探る調

査の中で、データサイエンスやデータ分析は、大学生にとって自分の生活とは関係しない遠い存在であると感じていること、高校までの数学に対する苦手意識を引きずっていることを示している。そして、データサイエンスは自分にとって関係があり自分にもできそうだと実感させるために、大学生にとって身近な社会問題を扱い、データ分析と問題解決を結び付けて考える意欲や発想力を育てることが重要であるとしている。また、竹内¹⁰⁾は、文系女子大学でのデータサイエンス教育の導入における活動結果と実施の際の課題について紹介している。そして、数理的な内容に戸惑う学生がいたり、嫌悪感を強く訴える学生もいるが、学んだことの利活用、なぜ学ぶのかの動機づけなどが重要であると述べている。さらに保田ら¹¹⁾は、私立文系大学におけるデータリテラシー教育科目の履修者に対して、意識および知識の変化について分析を実施している。その結果、興味および関心を高める面においては効果が認められるものの、知識の習得と理解を深める面においては課題が示され、理論的な部分の知識を学生にとって理解しやすい内容に改めるべく、授業設計をやり直し、授業改善を行いたいとする。

これらの先行研究は、文系大学生がデータサイエンスに対して距離感や苦手意識を持っており、学ぶための動機づけが重要であることを示している。

一方で内田¹²⁾は、文系大学生を対象としたデータサイエンスに関する意識調査を通して、苦手意識を持つ学生が多いものの、数理・データサイエンス・AIに対して、自分に必要な学習であるとポジティブに捉える学生も多いことを明らかにしている。そして、この分野は理解度が低い用語が多いことから、専門用語の解説を丁寧に行い、初学者にもわかりやすい説明が必要だとも述べている。

つまり先行研究によると、文系大学生はデータサイエンスに関する教育内容に対して苦手意識を持っていることから、初学者にもわかりやすい丁寧な説明と、学生が関心を持つ題材とすることで、学習意欲と理解度を向上させられるものと推察される。

2. 本研究の目的

2025年度までに、文理を問わず全ての大学・高専生が、数理・データサイエンス・AI教育プログラムのリテラシーレベルを身につけておくべき素養として位置付けられている。このような社会的背景において、苦手意識を持っていることが多い文系大学生を対象に、関心を持つ題材を用いて、わかりやすい丁寧な説明をする授業を実施し、学習意欲と理解度を向上させることは急務である。

そこで、「関心を持つ題材を用いて、わかりやすい丁寧な説明をする授業を実施し、学習意欲と理解度を向上させることは急務である。」

寧な説明をする授業」という部分に着目し、データサイエンスに関する文系大学生の理解度向上を目指した教授法を明らかにすることを本研究の目的とした。

3. 模擬授業の内容

(1) 模擬授業で取り扱う項目の選定

数理・データサイエンス・AI教育プログラムのリテラシーレベルのコア学修項目（必須項目）の中から、文系大学生が苦手意識を持っていると考えられる項目を抽出し検討した。その結果、基礎（データリテラシー）に含まれる標準偏差を模擬授業の内容とした。標準偏差を学習するためには、偏差、分散という理解度が低いと考えられる専門用語が登場することや、数式を使った説明や表を使った説明、表計算ソフトを使った説明など、様々な方法で説明できることから、本研究で実施する模擬授業の内容に相応しいと判断した。

(2) 教授法の選定

標準偏差を学習するための教授法として、数式、表、Excel（表計算ソフト）の3種類を選定した。数式を選定した理由は、従来から実施されている理系大学生を主な対象としたデータサイエンス教育では、数学の知識を前提として数式を活用する教授法で実施されているためである。表を選定した理由は、偏差や分散を求めて標準偏差に辿り着く過程を、表を用いて示すことで、わかりやすく説明できるためである。Excelを選定した理由は、業務の中で活用する際に、Excel等の表計算ソフトを用いて、搭載された関数で計算するためである。

(3) 基礎知識の確認

数式、表、Excelを用いて標準偏差を学習させるにあたり、事前に偏差と分散についての確認を実施した。題材は図-1に示すように、国語と英語のテストの点数とし、平均点がどちらも50点の場合、ばらつきが大きいのはどちらだろうかという身近で学生が興味を持つと考えられる内容にした。そして、各自の点数（値）と平均点（平

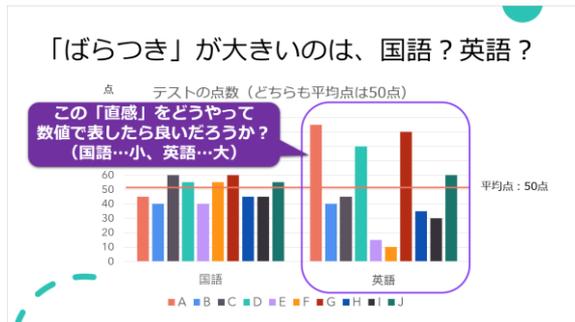


図-1 事前に実施した偏差と分散の確認（導入部分）

均）との差が偏差であり、偏差を全て足したら0になること、偏差の平均も0になることを確認した。次に、負の値を二乗すると正の値になる特徴を復習した上で、偏差を二乗した数値の平均値が分散だと確認した。

(4) 数式を活用した模擬授業

数式を活用した授業では、総計n個あるi番目のデータを x_i 、平均値をzとすると、分散 s^2 は以下の式となることを説明し、この式で国語と英語の分散を求めた。

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - z)^2$$

次に、単位を元のデータと揃えるために、平方根を計算したものが標準偏差であることを説明した。そのため、総計n個あるi番目のデータを x_i 、平均値をzとすると、標準偏差sは以下の式となることを説明し、図-2に示すように国語と英語の標準偏差を求めた。

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - z)^2}$$

計算結果から、国語の標準偏差は7.42、英語の標準偏差は28.64となり、国語よりも英語の方が、ばらつきが大きいことを数値によって示すことができると説明した。

(5) 表を活用した模擬授業

表を活用した授業では、偏差を二乗した数値の平均値が分散であるため、偏差と偏差の二乗を図-3に示す表に

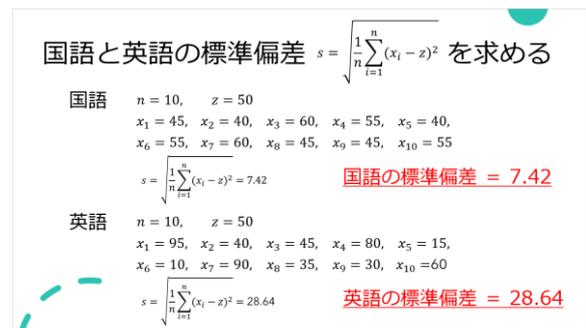


図-2 数式を使って標準偏差を計算する説明



図-3 表を使って標準偏差を計算する説明

整理し、国語と英語の分散を求めた。次に、単位を元のデータと揃えるために、平方根を計算したものが標準偏差であることを説明し、表に整理することで標準偏差を求めた。表に整理した計算結果から、国語（7.42）よりも英語（28.64）の方が、ばらつきが大きいことを数値によって示すことができると説明した。

(6) Excelを活用した模擬授業

Excelを活用した授業では、分散を求める関数がVAR.Pであることから、この関数を用いて分散を求めた。次に、単位を元のデータと揃えるために、平方根を計算したものが標準偏差であることと、標準偏差を求める関数がSTDEV.Pであることを説明した。そして図4に示すようにこの関数を用いて標準偏差を求めた。計算結果から、国語（7.42）よりも英語（28.64）の方が、ばらつきが大きいことを数値によって示すことができると説明した。

4. 文系大学生を対象とした調査

(1) 期日と対象

2024年5～6月に、S県に所在する文系の大学において、「情報倫理」を受講する203名（aクラス：68名、bクラス：71名、cクラス：64名）を対象に、授業の45分間を利用して実施した。

(2) 調査の目的

対象の学生は、主に国際経営学部の1年生（一部再履修者が含まれる）であり、これまでに大学でデータサイエンスや統計学に関する学習をしていない学生である。



図4 Excelを使って標準偏差を計算する説明

表-1 調査の流れ（aクラス）

	展開	時間
【基礎の確認】	偏差と分散についての確認	7分
【模擬授業①】	Excelを活用した模擬授業	8分
【調査①】	Excelを活用した模擬授業に対する意識調査	3分
【模擬授業②】	数式を活用した模擬授業	8分
【調査②】	数式を活用した模擬授業に対する意識調査	3分
【模擬授業③】	表を活用した模擬授業	8分
【調査③】	表を活用した模擬授業に対する意識調査	3分
【調査④】	各教授法に関する共通項目と属性調査	5分

このような学生を対象にデータサイエンスに関する模擬授業を行い、意識調査によって理解度向上を目指した教授法を明らかにすることを目的とした。

表-1に示すのはaクラスの調査の流れである。数式、表、Excelの模擬授業を実施する順番が、調査結果に影響を及ぼす可能性も否定できないため、クラスごとに順番を変更した。なお、bクラスは、数式→Excel→表、cクラスは、表→Excel→数式の順である。

(3) 調査の方法

調査項目は表-2に示す通りである。データサイエンスに関する既存の考え方などを確認する「データサイエンスに関する認識調査項目」と、理解のしやすさや対象者の属性を確認する「各教授法に関する共通項目と属性」、各教授法に対する理解度や評価を確認する「各教授法に関する調査項目」に分類される。例えば、表-1に示すaクラスの調査の流れでは、調査①でQ36～Q40、調査②でQ26～Q30、調査③でQ31～Q35、調査④でQ24～Q25を、それぞれMicrosoft Formsから制限時間以内に回答させる流れである。なお、Q1～Q23は事前に調査を実施した。

回答は特記の無い限り4段階のリッカートスケールとした。なお、Q8は自由記述、Q11は「小学校レベルの算数」、「中学校レベルの数学」、「高校レベルの数学」、「大学レベルの数学」、「大学での数学科のような専門的なレベルの数学」、「数学の知識は必要ない」の6択とした。Q24は「数式を使った授業」、「表を使った授業」、「Excelを使った授業」の3項目を理解できると思う順に並べることにした。Q25は「文系」、「理系」の2択である。

(4) 調査結果

Q1～Q40までの全ての項目に回答していたのは173名だった。この173名の回答について、Q1～Q7、Q9～Q10、Q12～Q23の平均・標準偏差を求めた結果を表-3に示す。

また、データサイエンスを学習するためには、どの程度の数学の知識が必要だと考えているかを尋ねたQ11の結果を表-4に示す。

そして、Q24の結果を表-5に、Q24の結果について1位を3点、2位を2点、3位を1点とし、人数を乗じて点数化した結果を表-6に示す。また、各教授法に関する調査項目であるQ26～Q40の結果を表-7に示す。

なお、対象者の出身高校（所属コース等）が文系か理系かを問うQ25の結果は、文系が133名（76.9%）、理系が40名（23.1%）だった。

(5) 分析および考察

Q1～Q3、Q5～Q7の結果が2.5未満であることから、データサイエンスがどのようなものなのか、学ぶ内容や所

表-2 質問項目

<データサイエンスに関する認識調査項目>

Q1: 「データサイエンス」という言葉を聞いたことはありますか?

Q2: 「データサイエンス」とは何か、理解していますか?

Q3: 「データサイエンス」に興味はありますか?

Q4: 「データサイエンス」を学ぶ必要があると思いますか?

Q5: 皆さんの所属学部で「データサイエンス」の授業が必修化されたことを知っていますか?

Q6: 「データサイエンス」の授業で、どのようなことを学ぶのか知っていますか?

Q7: 「データサイエンス」の授業で、どのようなことを学ぶのかイメージをすることはできますか?

Q8: 【自由記述】「データサイエンス」の授業では、どのようなことを学ぶと思いますか? (簡潔に)

Q9: 「データサイエンス」と数学は関係していると思いますか?

Q10: 「データサイエンス」と科学は関係していると思いますか?

Q11: 【6択問題】「データサイエンス」を学習するためには、どの程度の数学の知識が必要だと思いますか?

Q12: 「データサイエンス」を学ぶのは、難しいと思いますか?

Q13: 「データサイエンス」を学ぶのは、楽しいと思いますか?

Q14: 「データサイエンス」の授業は必修ですが、単位が取れるか不安ですか?

Q15: 数学は好きですか?

Q16: 数学は得意ですか?

Q17: 小学校レベルの算数は理解できていますか?

Q18: 中学校レベルの数学は理解できていますか?

Q19: 高校レベルの数学は理解できていますか?

Q20: データ分析をしたことはありますか?

Q21: データ分析は面白そうだと思いますか?

Q22: データ分析をやってみたいと思いますか?

Q23: ビジネスや社会問題を解決するために、文系でもデータサイエンスの知識が社会から求められていることを知っていますか?

<各教授法に関する共通項目と属性>

Q24: 【並べ替え問題】どの形式で「データサイエンス」の授業を実施すれば、あなたは内容を理解できると思いますか? 理解できると思う順に並べてください。

Q25: 【2択問題】出身高校(所属コース等)は文系ですか? 理系ですか? 近い方を選択してください。

<各教授法に関する調査項目>

Q26: 数式を使った授業は、わかりやすかったですか?

Q27: 数式を使った授業で、内容を理解できましたか?

Q28: 数式を使った授業で「データサイエンス」の授業を実施した場合、単位が取れるか不安ですか? (逆転項目)

Q29: 数式を使った授業で「データサイエンス」の授業を実施した場合、社会に出たときに役に立つ知識を身に付けることができると思いますか?

Q30: 数式を使った授業で「データサイエンス」の授業を実施した場合、授業で学んだことを将来に活かせると思いますか?

Q31: 表を使った授業は、わかりやすかったですか?

Q32: 表を使った授業で、内容を理解できましたか?

Q33: 表を使った授業で「データサイエンス」の授業を実施した場合、単位が取れるか不安ですか? (逆転項目)

Q34: 表を使った授業で「データサイエンス」の授業を実施した場合、社会に出たときに役に立つ知識を身に付けることができると思いますか?

Q35: 表を使った授業で「データサイエンス」の授業を実施した場合、授業で学んだことを将来に活かせると思いますか?

Q36: Excelを使った授業は、わかりやすかったですか?

Q37: Excelを使った授業で、内容を理解できましたか?

Q38: Excelを使った授業で「データサイエンス」の授業を実施した場合、単位が取れるか不安ですか? (逆転項目)

Q39: Excelを使った授業で「データサイエンス」の授業を実施した場合、社会に出たときに役に立つ知識を身に付けることができると思いますか?

Q40: Excelを使った授業で「データサイエンス」の授業を実施した場合、授業で学んだことを将来に活かせると思いますか?

表-3 Q1~Q7, Q9~Q10, Q12~Q23の結果 (n=173)

項目	平均	S.D.	項目	平均	S.D.	項目	平均	S.D.
Q1	2.35	1.20	Q9	3.23	0.78	Q17	3.62	0.65
Q2	1.54	0.79	Q10	3.24	0.86	Q18	3.25	0.75
Q3	2.30	0.89	Q12	3.39	0.75	Q19	2.45	0.90
Q4	2.95	0.87	Q13	2.54	0.92	Q20	2.42	1.11
Q5	1.45	0.87	Q14	3.25	0.90	Q21	2.58	1.01
Q6	1.50	0.77	Q15	2.03	1.07	Q22	2.66	1.01
Q7	2.14	0.91	Q16	1.97	1.02	Q23	2.21	1.15

表-4 Q11の結果 (n=173)

項目	人数(人)	割合(%)
小学校レベルの数学	8	4.6
中学校レベルの数学	55	31.8
高校レベルの数学	79	45.7
大学レベルの数学	17	9.8
大学での数学科のような専門的なレベルの数学	6	3.5
数学の知識は必要ない	8	4.6

表-5 Q24の結果 (n=173)

1位			2位			3位		
教授法	人数	割合(%)	教授法	人数	割合(%)	教授法	人数	割合(%)
数式	8	4.6	数式	19	11.0	数式	146	84.4
表	99	57.2	表	59	34.1	表	15	8.7
Excel	66	38.2	Excel	95	54.9	Excel	12	6.9

表-6 Q24を点数化した結果

教授法	点数化	点数
数式	$3 \times 8 + 2 \times 19 + 1 \times 146$	208
表	$3 \times 99 + 2 \times 59 + 1 \times 15$	430
Excel	$3 \times 66 + 2 \times 95 + 1 \times 12$	400

表-7 Q26~Q40の結果 (n=173)

数式			表			Excel		
項目	平均	S.D.	項目	平均	S.D.	項目	平均	S.D.
Q26	2.11	0.95	Q31	3.18	0.75	Q36	3.04	0.81
Q27	2.11	0.94	Q32	3.14	0.74	Q37	3.02	0.75
Q28	3.39	0.84	Q33	2.71	0.95	Q38	2.76	0.94
Q29	2.57	0.99	Q34	3.15	0.73	Q39	3.36	0.69
Q30	2.60	1.04	Q35	3.15	0.76	Q40	3.38	0.70

属している学部で必修化になったことを知らず、興味もあまりないことが読み取れる。しかし、Q4の結果から、多くの学生がどちらかと言えば学ぶ必要があると考えていることも推測できる。Q9~Q10より、データサイエンスは数学や科学と関係していると考えており、Q12, Q14の結果から、データサイエンスを学ぶのは難しく、単位を取得できるのか不安に思っていると考えられる。また、Q13の結果から、データサイエンスを学ぶことが楽しいかどうかは各々考えが分かれた。データサイエンスで学ぶ内容を問うQ8の自由記述では、データの分析やデータの仕組み、情報処理などの記述が多く見られた。

数学に関する項目では、データサイエンスを学習するためには、どの程度の数学の知識が必要だと考えているかを尋ねたQ11の結果から、高校レベルの数学の知識が必要だと考えていることが示された。また、Q15~Q16より、数学はあまり好きではなく、得意でもないことが読み取れる。Q17~Q19の結果から、中学校レベルの数学まではある程度理解できているが、高校レベルの数学になると半数以上が理解できていないと回答している。

データ分析に関する項目では、データ分析の経験の有

無を問うQ20の結果はばらけた。また、データ分析は面白いと思うかを問うQ21、やってみたいと思うかを問うQ22については、やや前向きであることが読み取れる。そして、Q23の結果から文系でもデータサイエンスの知識が社会から求められていることについて、あまり知られていないことが示された。

続いて、標準偏差を学習する模擬授業を受講した後の意識調査の結果を考察する。表-6に示すQ24で得られた順位付けを点数化した結果から、表を活用した授業が430点となり、最も内容を理解できると考えていることが明らかとなった。また、Excelを活用した授業も400点と高い評価が得られた。一方で、数式を活用した授業は208点であった。

次に、各教授法に関する5問ずつの調査項目（Q26～Q30, Q31～Q35, Q36～Q40）の結果を考察する。各項目の1～3問目（Q26～Q28, Q31～Q33, Q36～Q38）は、理解に関する項目のため「理解度スコア」とし、各項目の4～5問目（Q29～Q30, Q34～Q35, Q39～Q40）は、将来に向けた有用性に関する項目のため「有用性スコア」として詳細な分析をすることにした。各教授法の理解度スコアと有用性スコアの比較を表-8に、各教授法における有意差の確認（分散分析および多重比較検定）を表-9に示す。また、表-10に各教授法の理解度スコアにおける有意差の確認を、表-11に各教授法の有用性スコアに

おける有意差の確認を示す。

表-8に示す分析結果によると、Excelが全体集計および有用性スコアで平均値が最も高く、理解度スコアに関しては表の平均値が最も高かった。表-9に示す各教授法における有意差の確認では、分散分析を行った結果、 $F(2, 516) = 66.25, p < .001, \eta^2 = .20$ であり、条件間に有意差が認められた。そして多重比較検定（Fisher's LSD.）を実施したところ、数式と表、数式とExcelに有意差が認められた。表-10に示す各教授法の理解度スコアにおける有意差の確認では、分散分析を行った結果、 $F(2, 516) = 53.50, p < .001, \eta^2 = .17$ であり、条件間に有意差が認められた。そして多重比較検定（Fisher's LSD.）を実施したところ、数式と表、数式とExcelに有意差が認められた。表-11に示す各教授法の有用性スコアにおける有意差の確認では、分散分析を行った結果、 $F(2, 516) = 45.95, p < .001, \eta^2 = .15$ であり、条件間に有意差が認められた。そして多重比較検定（Fisher's LSD.）を実施したところ、全ての組み合わせ（数式と表、数式とExcel、表とExcel）に有意差が認められた。

このように、各教授法に関する調査項目からは、表とExcelを活用した授業の評価が高く、有用性スコアでは表よりもExcelの評価が有意に高い分析結果が示された。

さらに、対象者の属性が各教授法の結果と関連しているか否かを確認するために、データサイエンスへの興味（Q3）、数学の得意/不得意（Q16）、高校レベルの数学の理解（Q19）、データ分析は面白そうだ（Q21）、文系/理系（Q25）の5項目について、相関係数を算出した。その結果を表-12に示す。

表-12の結果によると、データサイエンスへの興味（Q3）は、表（全体）に、数学の得意/不得意（Q16）と高校レベルの数学の理解（Q19）は、数式（全体、理解度スコア、有用性スコア）に、やや相関関係がある（0.2以上0.4未満）と示された。

一方、データサイエンスへの興味（Q3）、数学の得意/不得意（Q16）、高校レベルの数学の理解（Q19）、データ分析は面白そうだ（Q21）、文系/理系（Q25）の5項目が、各教授法の結果に及ぼす影響を調査するた

表-8 各教授法の理解度/有用性スコア (n=173)

教授法	全体		理解度スコア		有用性スコア	
	平均	S.D.	平均	S.D.	平均	S.D.
数式	2.56	0.61	2.54	0.52	2.59	0.98
表	3.06	0.44	3.01	0.42	3.15	0.69
Excel	3.11	0.42	2.94	0.43	3.37	0.64

表-9 各教授法における有意差の確認

分散分析表					多重比較検定 (Fisher's LSD.)				
SV	SS	df	MS	F	水準1	水準2	統計量	p値	有意性
条件	32.82	2	16.41	66.25	数式	表	9.48	p<.001	**
誤差	127.82	516	0.25		数式	Excel	10.39	p<.001	**
全体	160.64	518		** p<.01	表	Excel	0.91	0.36	n.s.

** p<.01, * p<.05, n.s. p≥.05

表-10 各教授法の理解度スコアにおける有意差の確認

分散分析表					多重比較検定 (Fisher's LSD.)				
SV	SS	df	MS	F	水準1	水準2	統計量	p値	有意性
条件	22.43	2	11.22	53.50	数式	表	9.55	p<.001	**
誤差	108.20	516	0.21		数式	Excel	8.22	p<.001	**
全体	130.63	518		** p<.01	表	Excel	1.33	0.18	n.s.

** p<.01, * p<.05, n.s. p≥.05

表-11 各教授法の有用性スコアにおける有意差の確認

分散分析表					多重比較検定 (Fisher's LSD.)				
SV	SS	df	MS	F	水準1	水準2	統計量	p値	有意性
条件	56.48	2	28.24	45.95	数式	表	6.69	p<.001	**
誤差	317.12	516	0.61		数式	Excel	9.29	p<.001	**
全体	373.59	518		** p<.01	表	Excel	2.61	0.01	**

** p<.01, * p<.05

表-12 対象者の属性と各教授法との相関係数

	全体			理解度			有用性		
	数式	表	Excel	数式	表	Excel	数式	表	Excel
Q3: データサイエンスへの興味	0.11	0.22	0.03	0.15	0.17	0.03	0.05	0.19	-0.02
Q16: 数学の得意/不得意	0.26	0.16	0.09	0.21	0.08	0.02	0.23	0.19	0.15
Q19: 高校レベルの数学の理解	0.32	0.15	0.03	0.24	0.06	-0.09	0.31	0.19	0.16
Q21: データ分析は面白そうだ	0.12	0.09	0.17	0.11	0.03	0.15	0.11	0.13	0.13
Q25: 文系/理系	0.17	0.08	0.07	0.13	0.00	0.03	0.17	0.13	0.11

表-13 回帰分析で得られたp値

	全体			理解度			有用性		
	数式	表	Excel	数式	表	Excel	数式	表	Excel
Q3: データサイエンスへの興味	0.14	0.00**	0.70	0.05*	0.02*	0.74	0.48	0.01*	0.79
Q16: 数学の得意/不得意	0.00**	0.03*	0.22	0.01**	0.29	0.76	0.00**	0.01*	0.05
Q19: 高校レベルの数学の理解	0.00**	0.05	0.67	0.00**	0.46	0.23	0.00**	0.01*	0.03*
Q21: データ分析は面白そうだ	0.10	0.23	0.03*	0.15	0.74	0.06	0.16	0.10	0.08
Q25: 文系/理系	0.02*	0.30	0.38	0.09	0.99	0.66	0.03*	0.10	0.17

** $p < .01$, * $p < .05$

表-14 回帰分析で得られた決定係数

	全体			理解度			有用性		
	数式	表	Excel	数式	表	Excel	数式	表	Excel
Q3: データサイエンスへの興味	0.01	0.05	0.00	0.02	0.03	0.00	0.00	0.04	0.00
Q16: 数学の得意/不得意	0.07	0.03	0.01	0.04	0.01	0.00	0.05	0.04	0.02
Q19: 高校レベルの数学の理解	0.10	0.02	0.00	0.06	0.00	0.01	0.09	0.03	0.03
Q21: データ分析は面白そうだ	0.02	0.01	0.03	0.01	0.00	0.02	0.01	0.02	0.02
Q25: 文系/理系	0.03	0.01	0.00	0.02	0.00	0.00	0.03	0.02	0.01

めに回帰分析を実施した。この回帰分析から得られたp値(有意F)を表-13に、決定係数(R²)を表-14に示す。

表-13の結果から、データサイエンスへの興味(Q3)と、数式(理解度)および(表全体, 理解度スコア, 有用性スコア), 数学の得意/不得意(Q16)と、数式(全体, 理解度スコア, 有用性スコア)および表(全体, 有用性スコア), 高校レベルの数学の理解(Q19)と数式(全体, 理解度スコア, 有用性スコア)と表(有用性スコア)およびExcel(有用性スコア), データ分析は面白そうだ(Q21)とExcel(全体), 文系/理系(Q25)と数式(全体, 有用性スコア)で有意な結果($p < .05$)となった。

しかし、表-14の結果から、決定係数の値が最も高い高校レベルの数学の理解(Q19)と表(全体)でも0.10と小さく、全ての組み合わせにおいて、説明変数が目的変数を十分に説明できる結果は得られなかった。

5. 結論

文系大学生にデータサイエンス教育を実施する場合、データサイエンスに関する教育内容に対して苦手意識を持っていることから、初学者にもわかりやすい丁寧な説明と、学生が興味を持つ題材とすることで、学習意欲と理解度を向上させられるものと推察されることを先行研究から知得した。

そこで本研究では、「興味を持つ題材を用いて、わかりやすい丁寧な説明をする授業」という部分に着目し、データサイエンスに関する文系大学生の理解度向上を目指した教授法を明らかにすることを目的とした。そして、標準偏差をテーマに、数式、表、Excelの3種類の教授法を用いて理解度と有用性に関する調査を実施した。以下に本研究で得られた知見をまとめる。

- ① データサイエンスに関する認識調査項目の結果から、データサイエンスに関する知識はほとんど無いことが示された。また、数学はあまり好きではなく、得意でもなく、高校レベルの数学になると半数以上が理解できていないと回答していた。このように、レディネスはほぼ無い状態であることが明らかとなった。
 - ② 標準偏差を学習する模擬授業を受講した後に実施した意識調査のQ24で、3つの教授法を理解できると思う順に並べ替えさせたところ、1位が表を活用した授業、2位がExcelを活用した授業、3位が数式を活用した授業の順となった。また、点数化したところ、表を活用した授業が430点となり、最も内容を理解できていると考えていることが明らかとなった。Excelを活用した授業も400点と高い評価が得られた一方で、数式を活用した授業は208点であった。このように、表を活用した授業が最も内容を理解できると評価された。
 - ③ 理解に関する項目である「理解度スコア」と、将来に向けた有用性に関する項目である「有用性スコア」に分け、詳細な分析を実施した。その結果、Excelを活用した授業が全体集計および有用性スコアで平均値が最も高く、理解度スコアに関しては、表を活用した授業の平均値が最も高かった。また、分散分析および多重比較検定を実施した結果、有用性スコアについては、表よりもExcelの方が有意に高く、調査対象の文系大学生は、Excelを活用した授業の方が有用であると回答していることが明らかとなった。
 - ④ 対象者の属性が各教授法の結果と関連しているか否かを確認した結果、データサイエンスへの興味(Q3)は、表(全体)に、数学の得意/不得意(Q16)と高校レベルの数学の理解(Q19)は、数式(全体, 理解度スコア, 有用性スコア)に、やや相関関係があると示された。
 - ⑤ 対象者の属性が各教授法の結果に及ぼす影響を調査するために回帰分析を実施した結果、全ての組み合わせにおいて、説明変数が目的変数を十分に説明できる結果は得られなかった。
- これらの結果から総合評価を実施する。まず、調査対象の文系大学生が、3つの教授法を理解できると思う順

に自ら並べ替えをした結果、表を活用した授業が最も内容を理解できると評価された。一方で、各教授法に関する調査項目からは、表とExcelを活用した授業の評価が高く、有用性スコアでは表よりもExcelの評価が有意に高いという分析結果が示された。これらの結果を総合的に評価すると、Excelを活用した授業が最も効果的であると結論付けられる。表を活用した授業の評価も高いが、特にExcelは高い有用性スコアを示し、幅広い学生に対して一貫して効果的であるため、データサイエンス教育において最適な教授法であると考えられる。

しかし、データサイエンス教育でExcelのみを活用するのではなく、表を活用したり、一部で数式を活用したりするなど、学習内容や対象者に合わせた授業を実施することが重要であると考えられる。

参考文献

- [1] 独立行政法人情報処理推進機構 (IPA), データサイエンティスト協会, 「データサイエンティストのためのスキルチェックリスト/タスクリスト概説」, 2020, (2024年7月12日取得, <https://www.ipa.go.jp/jinzai/skill-standard/plus-it-ui/itssplus/ps6vr7000001ity-att/000083733.pdf>).
- [2] 内閣府, 「第5期科学技術基本計画」, 2016, (2024年7月12日取得, <https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>).
- [3] 内閣府, 「人間中心のAI社会原則」, 2019, (2024年7月19日取得, <https://www8.cao.go.jp/cstp/aigensoku.pdf>).
- [4] 内閣府, 「AI戦略2019～人・産業・地域・政府全てにAI～」, 2019, (2024年7月16日取得, <https://www8.cao.go.jp/cstp/ai/aistrategy2019.pdf>).
- [5] 数理・データサイエンス・AI教育強化拠点コンソーシアム, 「数理・データサイエンス・AI教育強化拠点コンソーシアムとは」, (2024年7月19日取得, <http://www.mi.u-tokyo.ac.jp/consortium/overview.html>).
- [6] 内閣府, 「「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)」の創設について」, (2024年7月19日取得, <https://www8.cao.go.jp/cstp/ai/suuri/ninteisousetu.pdf>).
- [7] 文部科学省, 「数理・データサイエンス・AI教育プログラム認定制度」, (2024年7月19日取得, https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm).
- [8] 数理・データサイエンス・AI教育強化拠点コンソーシアム, 「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～」, 2024, (2024年7月19日取得, http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy_20240222.pdf).
- [9] 松尾由美, 玉田和恵, 「文系・私立大学におけるデータサイエンス教育の課題」, 江戸川大学紀要, 第31号, pp.249-255, 2021.
- [10] 竹内光悦, 「文系女子大学におけるデータサイエンス教育の導入と実施の課題」, 実践女子大学人間社会学部紀要, 第18集, pp.111-117, 2022.
- [11] 保田洋, 宇佐美美紀子, 「私立文系大学のデータサイエンス教育の円滑な導入に向けた学生の意識の変化についての分析」, 情報知識学会誌, Vol.33, No.2, pp.137-141, 2023.
- [12] 内田いづみ, 「文系学生に対するAI・データサイエンス教育の効果的アプローチの考察」, 駿河台大学論叢, 第65号, pp.85-96, 2023.

2024年?月?日 受稿

2024年?月?日 受理

COMPARATIVE ANALYSIS OF TEACHING METHODS TO IMPROVE UNDERSTANDING OF DATA SCIENCE EDUCATION AMONG HUMANITIES STUDENTS

Rena KAWAI, Taiga ITO and Saki ITABASHI

In Japan's "Mathematics, Data Science, and AI Education Program," literacy level is positioned as "essential knowledge that all university students should acquire," requiring even humanities students to learn data science. However, humanities students often have a strong aversion to mathematics, making it difficult for them to understand traditional teaching methods that utilize mathematical formulas.

Therefore, this study conducted three types of mock classes on the topic of standard deviation using formulas, tables, and Excel, targeting humanities students to investigate which teaching method is most appropriate. The results showed that the class using tables received the highest evaluation. Furthermore, the class using Excel was considered to provide knowledge that would be useful in society and applicable in the future. Overall, the class using Excel was shown to be the most suitable teaching method for humanities students.