

学校教員を対象としたオンライン研修「教育データサイエンスセミナー」の 試行的実践と評価

－教員養成課程における AI・データサイエンス科目の開発に向けて－

Trial Practice and Evaluation of the Online Training 'Educational Data Science Seminar' for School Teachers: Toward the Development of AI and Data Science Subjects in Teacher Training Courses

清水 優 菜* 掛 川 淳 一** 森 山 潤***
SHIMIZU Yuno KAKEGAWA Junichi MORIYAMA Jun

兵庫教育大学では、2020 年度より教員養成・研修高度化センターに教育データサイエンスチームが設置され、学部学生全員が履修する AI・データサイエンス科目として「AI・データサイエンス基礎」と「教育データサイエンス」が構想された。この 2 科目の試行的実践として、現職の学校教員を対象に「教育データサイエンスセミナー」を実施した。その結果、セミナーの満足度は高く、かつ難易度は適切な水準であることが示された。さらに、セミナーはデータサイエンスに関する知識習得の重要性と教育データサイエンスを学習・活用することにメリットがあることを認識させるものであった。以上から、構想した 2 科目は、教員養成における AI・データサイエンス科目として適切なものと考えられる。他方、内容の改善点として、データ収集の方法、AI に関する知識の習得、教育における AI の活用を拡充することが示された。これらの改善点に即して、構想した 2 科目を修正・実施・評価することが今後求められる。

キーワード：兵庫教育大学、教育データサイエンス、教員研修、オンライン、試行的実践

Key words : hyogo university of teacher education, educational data science, teacher training, online, trial practice

1. 問題と目的

IoT, ビッグデータ, 人工知能 (以下, AI) をコアとする第四次産業革命では、我々人間の営為を逐一データ化し、得られたデータを分析することで、多様かつ困難な問題を解決することが目指されている。その中、データサイエンスは万人が身につけるべき基礎的素養 (リテラシー) の 1 つと定位されつつある。小学校教育では、プログラミング教育が必修になるとともに、算数において「D データの活用」領域が新設された (文部科学省, 2017a, 2017b)。高等学校教育では、共通教科情報にデータサイエンスに関する内容が新設・拡充されるとともに、数学 B において正規分布を利用した区間推定と仮説検定の考え方が導入される (文部科学省, 2018a, 2018b)。高等教育では、滋賀大学や横浜市立大学をはじめとしてデータサイエンス学部が創設されるとともに、国の目標として「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」することが掲げられている (内閣府, 2019)。

さらに、近年では、多様な子供たちの「誰一人取り残すことのない、公正に個別最適化された学び」を実現するために、学校教員が ICT 環境を基盤とした先端技術や教育ビッグデータを活用することが求められている

(文部科学省, 2019)。具体的には、学校教員は先端技術や教育ビッグデータを活用することで、以下 4 点が期待されている。その 1 に、個々の子供の状況を客観的かつ継続的に把握するなどの「個別に最適で効果的な学びや支援」である。その 2 に、遠隔技術を活用した大学や海外との連携授業などの「学びにおける時間・距離などの制約を取り払う」ことである。その 3 に、遠隔技術を活用した場所に制約を受けない教員研修や採点業務などの「校務の効率化」である。その 4 に、学習履歴、行動等の様々なビッグデータ分析による「経験知」の可視化などの「学びの知見の共有や生成」である。また、国立教育政策研究所では、教育政策においてデータを効果的かつ効率的に取得・分析・活用して、効果的な政策を立案・実施するために、我が国の教育データ分析・研究、成果共有の拠点として、2021 年度 10 月に「教育データサイエンスセンター」が設置された (国立教育政策研究所, n.d.)。このセンターでは、教育データや取組を共有するための基盤整備、教育データ分析・研究の推進、国や自治体における教育データ分析・研究の支援が取り組まれており、教育ビッグデータの活用が現在進行形で急速に進んでいる。

以上の動向を踏まえると、先端技術やデータサイエンスを指導・活用できる教員の養成・研修は、教員養成大

* 兵庫教育大学先端教職課程カリキュラム開発センター 助教

令和 4 年 7 月 14 日受理

** 兵庫教育大学大学院学校教育研究科人間発達教育専攻 生活・健康・情報系教育コース 准教授

*** 兵庫教育大学大学院学校教育研究科人間発達教育専攻 生活・健康・情報系教育コース 教授

表 1. 育成したい資質・能力

A.教育データを活用する力
A-1:教育データを収集する力
A-2:教育データを可視化する力
A-3:教育データを分析する力
A-4:教育データを解釈する力
B.AI 等を用いた多様な EdTech を活用する力
B-1:AI の特性に対する基本的な理解
B-2:AI を用いた EdTech を活用する力
B-3:EdTech と直接指導とを効果的に組み合わせた指導を展開できる力
C.教育データサイエンスの展望
C-1:教育ビッグデータの活用
C-2:学習分析(Learning Analytics)の理解
C-3:LMS の活用
C-4:e ポートフォリオ 等

学における主たる喫緊の課題である。その中で、兵庫教育大学では、2020 年度より教員養成・研修高度化センターに、本稿筆者らにより構成される教育データサイエンスチームが設置され、教員養成における AI・データサイエンス教育の在り方、方法、内容についての検討が進められてきた。

2020 年度は、データサイエンス教育に関する国内外の取り組みを調査・整理し、Education Data Science Master's Program (Graduate School of Education, Stanford University, n.d.) や「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」(数理・データサイエンス教育強化拠点コンソーシアム, 2020) などを参考にして、教育データサイエンスの位置付けと育成したい資質・能力を次のように規定した。教育データサイエンスの位置付けについては、「数理・データサイエンスリテラシーを基盤に、学校教育における職能として必要な内容知、方法知、問題解決能力を扱うもの」と規定した。育成したい資質・能力については、データを収集する力などの「エビデンスに基づく教育改善の力」、AI の特性に対する基本的な理解などの「AI 等を用いた多様な EdTech を活用する力」、教育ビッグデータの活用などの「教育データサイエンスの展望」の 3 要素から構成されるものとした (表 1)。

2021 年度は、規定した教育データサイエンスの位置付けと育成したい資質・能力に基づき、教員養成における AI・データサイエンス科目として学部学生全員が履修する「AI・データサイエンス基礎」と「教育データサイエンス」を構想した (表 2, 3)。まず、「AI・データサイエンス基礎」は、「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」(数理・データサイエンス教育強化拠点コンソーシアム, 2020) に準拠するものとした。第 1 回から第 12 回の内容は、データと AI の概要に関するもので、表 1 で示した資質・能力のうち「AI 等を用いた多様な EdTech を活用する力」の育成を目指す。第 13 回から第 15 回の内容は、実際にデータを扱うもので、「エビデンスに基づく教育改善の力」の育成を目指す。次に、「教育データサイエンス」は、「AI・データサイエンス基礎」を通して育成された資質・

能力を基礎として、「エビデンスに基づく教育改善の力」と「教育データサイエンスの展望」の育成を目指すものとした。第 1 回と第 15 回の内容は、教育ビッグデータや学習分析などの概論に関するもので、「教育データサイエンスの展望」の育成を目指す。第 2 回から第 14 回の内容は、記述統計と推測統計の基礎、ならびに統計的なデータ分析の演習に関するもので、「AI・データサイエンス基礎」で育成された「エビデンスに基づく教育改善の力」をより確固たるものにすることを目指す。

「AI・データサイエンス基礎」と「教育データサイエンス」の構想はしたもの、これらに類する科目は、兵庫教育大学の学校教育課程の必修科目としてこれまで行われていない。それゆえ、構想した科目の内容が学部学生、ひいては教員養成・研修にとって、どの程度適切なものか定かではない状態にあった。そこで、教育データサイエンスチームは、「AI・データサイエンス基礎」と「教育データサイエンス」の在り方を検討するために、現職の学校教員を対象に「教育データサイエンスセミナー」を実施した。セミナーでは、「AI・データサイエンス基礎」と「教育データサイエンス」の内容を網羅的に取り上げ、セミナー終了後に参加者から評価を得た。得られた評価を詳細に分析することで、「AI・データサイエンス基礎」と「教育データサイエンス」の実施に向けて、有益な知見を得ることが期待できよう。

以上を踏まえ、本稿は、「教育データサイエンスセミナー」の実践と評価を通して、「AI・データサイエンス基礎」と「教育データサイエンス」の在り方を検討することを目的とする。

2. 「教育データサイエンスセミナー」の実践

2.1. セミナーの概要

本セミナーは、「AI・データサイエンス基礎」と「教育データサイエンス」の試行的実践として、統計解析や機械学習等に関する学習経験のない現職の学校教員を対象とした。統計解析や機械学習などの初心者を対象とすることを鑑み、本セミナーでは教育データサイエンスの基礎的な内容やその具体を数式ではなく図的表現等により説明することに重きを置いた。

表 2. 「AI・データサイエンス基礎」のシラバス案

第1回	ガイダンス
第2回	社会で起きている変化
第3回	社会で活用されているデータ
第4回	データ・AI の活用領域
第5回	データ・AI 利活用の技術
第6回	データ・AI の活用現場と最新動向
第7回	データ・AI 活用の倫理(1)：ELSI, 個人情報保護
第8回	データ・AI 活用の倫理(2)：データ倫理
第9回	データ・AI 活用と社会の在り方(1)：AI 社会原則, データバイアス
第10回	データ・AI 活用と社会の在り方(2)：アルゴリズムバイアス, 責任論
第11回	データ・AI 活用の留意点：負の事例, 情報セキュリティ
第12回	AI と社会の発展
第13回	データリテラシー(1)：データを読む
第14回	データリテラシー(2)：データを説明する
第15回	データリテラシー(3)：データを扱う, まとめ

表 3. 「教育データサイエンス」のシラバス案

第1回	ガイダンス, 教育におけるデータの活用, 実務のイメージ
第2回	データの種類に応じた要約と可視化
第3回	統計解析の基礎理論(1)：確率, 期待値など
第4回	統計解析の基礎理論(2)：正規分布とその他の確率分布
第5回	推測統計の考え方
第6回	不偏推定量(標準誤差, 不偏分散, 標本標準偏差)
第7回	区間推定
第8回	統計的仮説検定の考え方
第9回	カイ 2 乗検定 [演習:児童生徒対象の意識調査の分析]
第10回	直接確率計算法, マクネマー検定 [演習:児童生徒の授業前後の意識の変容の分析]
第11回	t 検定 [演習:児童生徒の学力調査の分析]
第12回	対応あり t 検定, t 検定の注意点 [演習:児童生徒の授業前後の学力の変容の分析]
第13回	相関係数 [演習:児童生徒の授業に対する意識の分析]
第14回	総合演習 [演習:第 10 回から第 13 回に関わる応用]
第15回	エビデンスに基づく教育改善の展望

セミナーへの参加者を募るために、兵庫教育大学公式 Web サイトやイベント情報サイト「SENSEI イベントポータル」等への掲載、兵庫県内の連携協定締結教育委員会への案内チラシの送付、修了生・卒業生等連絡センターから修了生・卒業生へのメール配信を行った。その結果、小学校、中学校、高等学校の教員（一部に大学教員を含む）計 72 名から参加が申し込まれた。この 72 名のうち、第 1 回と第 2 回両方のセミナーへの参加申し込みがあったのは 65 名であった。

セミナーは、2021 年 8 月 1 日と 8 月 20 日の全 2 回、約 3 時間ずつ行われた。セミナーの概要を表 4 に記す。第 1 回セミナーは「教育データサイエンス事始め」と題して、教育データサイエンスの概要、記述統計の基礎、教育における AI の活用に関する講義・演習が行われた。第 2 回セミナーは「教育におけるデータの活用と可能性」と題して、統計的なデータ分析の概要、教育データの読み取り方、教育データサイエンスの可能性に関する講義・演習が行われた。次節では、セミナー各回の展開について、その詳細を示す。

2.2. 第 1 回セミナーの展開

第 1 回セミナーでは、「教育データサイエンス：全体

像を掴む」,「教育におけるデータ活用：記述統計の基礎」,「教育における AI の活用」の 3 講座が約 1 時間ずつ実施された。全講座終了後にセミナーの内容に関する質疑が行われた。以下では、実施された講座順に、その詳細を示す。

2.2.1. 教育データサイエンス：全体像を掴む

第 3 筆者を講師として、教育データサイエンスの概要について講義が行われた。具体的な展開は、次の 5 つからなる。その 1 に、データの定義と活用についてである。データを「事象を特定の視点から切り出し指標化した数値や資料」とした上で、人間は古くからデータを問題解決の手段として活用してきたことを事例により説明した。そして、コンピュータやインターネットの進展により、統計解析と AI からなるデータサイエンスの重要性が高まっていることを示した。その 2 に、データへのアプローチ方法についてである。データのアプローチとして統計解析と AI の 2 つを提示し、それぞれの概要を説明した。その 3 に、教育におけるデータの活用の意義と具体についてである。教育におけるデータ活用には教師、政策決定者、学習者という観点があること、ならびにデータ活用の流れについて説明した。その 4 に、こ

表 4. セミナーの概要

	第1回セミナー	第2回セミナー
日時	2021年8月1日(日) 13:10~16:20	2021年8月20日(金) 13:10~16:20
テーマ	教育データサイエンス事始め	教育におけるデータの活用と可能性
講義1	教育データサイエンス：全体を掴む (1) データの定義と活用 (2) データへのアプローチ方法 (3) 教育におけるデータの活用の意義と具体 (4) これからの教員に求められる資質・能力 (5) 実務のイメージ	統計的なデータ分析：全体像を掴む (1) 記述統計と推測統計 (2) 統計的推定 (3) 統計的仮説検定
講義2	教育におけるデータの活用：記述統計の基礎 (1) 統計学の枠組み (2) データの種類と形式 (3) データの代表値 (4) データの散布度 (5) 標準化得点と偏差値 (6) 正規分布 (7) データの可視化	教育データの読み取り方 (1) 関連の概念 (2) 全国学力・学習状況調査の読み取り方 (3) OECDによるPISA調査の読み取り方
講義3	教育におけるAIの活用 (1) AIの定義と概要 (2) 機械学習の概要 (3) ニューラルネットワークとディープラーニング (4) 生活での実装・応用例 (5) AI搭載型EdTechの事例	教育データサイエンスの可能性 (1) 教育ビッグデータ (2) 学習分析の概要と具体例 (3) 日本の小・中学校における学習履歴の利活用 (4) AI活用などの研究動向と将来展望

れからの教員に求められる資質・能力についてである。ここでは、表1で先述した育成したい資質・能力について説明した。その5に、実務のイメージについてである。ここでは、校内で実施するさまざまなアンケート調査の実施と分析などの教育データの活用、ならびにAI搭載のアダプティブなアプリを活用した個別最適化学習などのAIを用いた多様なEdTech活用の概略を説明した。

2.2.2. 教育におけるデータ活用：記述統計の基礎

第1筆者を講師として、記述統計の基礎について講義・演習が行われた。具体的な展開は、次の7つからなる。その1に、統計学の枠組みについてである。統計学は記述統計と推測統計からなること、ならびにこれらの具体例を説明した。その2に、データの種類と形式についてである。ここでは、名義データ、順序データ、間隔データ、比例データというデータの種類の種類とこれらの性質、行列によるデータの表現を説明した。その3に、データの代表値についてである。平均値、中央値、最頻値の概要を説明した上で、これらの特徴を比較する演習が行われた。

その4に、データの散布度についてである。四分位数、偏差、分散、標準偏差の概要を説明した上で、これらの概念を確認する演習が行われた。その5に、標準化得点と偏差値についてである。ここでは、標準化得点の概要を説明した上で、偏差値との関係を示した。その6に、正規分布についてである。正規分布の特徴と偏差値や教育評価との関係を説明した。その7に、データの可視化についてである。データを可視化する方法として、棒グラフ、折れ線グラフ、円グラフ、帯グラフ、積み上げ棒グラフ、レーダーチャート、度数分布表、ヒストグラム、箱ひげ図の特徴と読み取り方を説明した。

2.2.3. 教育におけるAIの活用

第2筆者を講師として、教育におけるAIの活用について講義が行われた。具体的な展開は、次の5つからなる。その1に、AIの定義と概要についてである。AIの定義、代表的な研究テーマ、AI研究のロードマップ、AIの最新動向を説明した。その2に、機械学習の概要についてである。機械学習の定義と応用例を提示した上

データって何?

- データ=事象を特定の視点から切り出し指標化⇒数値, 資料

体つき → 🧐「身長」 → 167.5cm

お菓子 → 🧐「カロリー」 → 300kcal

事象
(様々な要素が複合的に絡み合っている)

視点
(測定可能な変数として捉える)

データ
(量的・質的な値が得られる)

図1. 「教育データサイエンス：全体像を掴む」の説明スライド例

例題1における「9000万円」の解釈と代表値

200 300 400 400 400 500 600 800 9000 [万円]

- 9000万円を除外したときの平均値は450万円
➢450万円は代表値として許容可能な値では?
- 外れ値(異常値): データの全体的な位置から極端に離れた値
➢9000万円は外れ値と判断できる!
- 平均値はデータの重心であるため、外れ値に弱い!
➢だからと言って平均値が劣ると言うことではない。
➢平均値と中央値が大きく異なる場合は両方報告する!

図2. 「教育におけるデータ活用：記述統計の基礎」の演習に関する解説例

で、教師あり学習、教師なし学習、強化学習を取り上げ、これらのメカニズムを説明した。その3に、ニューラルネットワークとディープラーニングについてである。ニューラルネットワークとディープラーニングの定義を提示した上で、これらのメカニズムを説明した。その4に、生活での実装・応用例についてである。スマート家電などのAIの応用例、ならびにドライバーモニタリングなどのディープラーニングの応用例を説明した。その5に、AI搭載型EdTechの事例についてである。文部科学省の「新時代の学びにおける先端技術導入実証研究事業」で示された学習活動の可視化・行動分析、つまり分析、経済産業省の「EdTech ライブラリー」で示された適応型(Adaptive)教材の具体例を説明した。

機械学習の種類			
方法	定義	代表的な方法	応用例
教師あり学習 (Supervised Learning)	「教師データ」(入力と結果や正解にあたる出力のデータ)に基づき入力を変換する関数を学習	回帰分析、分類(近傍法、決定木、ナイーブベイズ、SVMなど)	故障診断・欠陥検査、顔認識、物体検出、手書き文字分類、売上予測、需要予測、人工予測、株価予測、降水量予測、不正検知など
教師なし学習 (Unsupervised Learning)	入力のみデータに基づき、入力の特性(パターン、構造)を学習	主成分分析、階層併合的クラスターリング、k平均法、アソシエーション分析、ソーシャルネットワーク分析など	データのグループ分け・要約・可視化(クラスターリング、次元圧縮)など
強化学習 (Reinforcement Learning)	入力と、出力に対する報酬(評価)のデータに基づき、入力を出力に変換する関数を学習	TD学習(Q学習など)など	ロボット制御(ロボットの歩行など)、ウェブ広告選択、マーケティング、ゲーム戦略獲得(将棋、囲碁など)、税金徴収の戦略獲得など

図3. 「教育におけるAIの活用」の説明スライド例

2.3. 第2回セミナーの展開

第2回セミナーでは、「統計的なデータ分析：全体像を掴む」、「教育データの読み取り方」、「教育データサイエンスの可能性」の3講座が約1時間ずつ実施された。第1回セミナーと同様に、全講座終了後にセミナーの内容に関する質疑が行われた。以下では、実施された講座順に、その詳細を示す。

2.3.1. 統計的なデータ分析：全体像を掴む

第3筆者を講師として、統計的なデータ分析の概要について講義・演習が行われた。具体的な展開は、次の3つからなる。その1に、記述統計と推測統計についてである。記述統計と推測統計の概要、全数調査と標本調査、母集団と標本を説明した。その2に、統計的推定についてである。点推定と区間推定の概要と具体例、母平均の

検定に使えるエクセル関数

- 下記の関数で、二つの平均値の差の検定ができる。
- ・t検定
=T.TEST(範囲1, 範囲2, 尾部, 検定の種類)

※範囲1に事前のデータ
※範囲2に事後のデータ
※尾部 1:片側検定 2:両側検定
※検定の種類 1:対応あり
2:対応なし, 等分散
3:対応なし, 等分散でない(ウェルチの検定)

演示

授業前後の平均値の比較 =T.TEST(A1:A20,B1:B20,2,1)
→ p値が出力される。p<0.05であれば有意。

図4. 「統計的なデータ分析：全体像を掴む」の演習例

区間推定を説明した上で、Excelにて母平均の区間推定を行う方法を演習した。その3に、統計的仮説検定についてである。帰無仮説、対立仮説、p値、統計的有意性を説明した。その上で、具体的な検定としてカイ2乗検定、マクネマー検定、t検定、分散分析の概要を説明し、Excelにてt検定を行う方法を演習した。演習後には、統計的仮説検定の非対称性(帰無仮説が成り立つ仮定が成り立たないことを検証しているが、対立仮説が成り立つことを積極的に検証していないこと)、ならびに第1, 2種の過誤を説明した。

2.3.2. 教育データの読み取り方

第1筆者を講師として、関連の概念と教育データの読み取り方について講義が行われた。具体的な展開は、次の3つからなる。その1に、関連の概念についてである。関連の定義、散布図、共分散、相関係数、相関係数の性質、散布図と相関係数の対応関係、疑似相関、共変量、相関関係と因果関係の違いを説明した。その2に、全国学力・学習状況調査の読み取り方についてである。全国学力・学習状況調査の概要とPISA調査との相違、古典的テスト理論と項目反応理論の概要を説明した。その上で、全国学力・学習状況調査の個人と集団の結果の読み取り方を説明した。その3に、OECDによるPISA調査の読み取り方についてである。PISA調査の概要を提示した上で、平均値と相関に基づくPISA調査の結果の読み取り方を説明した。

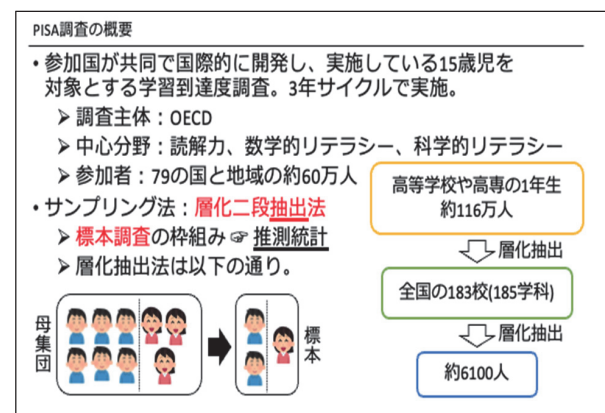


図5. 「教育データの読み取り方」の説明スライド例

2.3.3. 教育データサイエンスの可能性

第2筆者を講師として、教育データサイエンスの動向と展望について講義が行われた。具体的な展開は、次の4つからなる。その1に、教育ビッグデータについてである。ビッグデータの定義を示した上で、学校現場で蓄積される教育・学習ビッグデータの種類と活用事例を説明した。その2に、学習分析の概要と具体例についてである。学習分析の定義とプロセス、教育データマイニング、LMS (Learning Management System)、学習分析の研究例と教育改善例、マルチモーダル学習分析を説明した。その3に、日本の小・中学校における学習履歴の利活用についてである。ペンストロック情報の利用、ならびにテキストへのマーカー使用情報の利用を具体例として説明した。その4に、AI活用などの研究動向

と将来展望についてである。人工知能学会誌「人工知能」の特集記事を提示した上で、教育・学習における AI 活用の研究の動向、ならびに今後の展望を説明した。

教育・学習におけるAI活用の研究 (機械学習応用のものに限定)

- ・「自己説明」の自動評価 (Panaite et al. 2018) †
- ・学習中の学習者の映像からの注意散漫状態の検出 (Stewart et al. 2017) †
- ・学習行動により学習者をグルーピング (Zhang et al. 2017) †
- ・姿勢データに基づくフラストレーション検出 (Henderson et al. 2019) †
- ・オンライン試験における替え玉検出 (安田・小方2021)
- ・映像を用いた園児の関心の可視化 (山田ら2021)
- ・etc.
- ・機械学習手法間の性能比較 (Shores et al. 2011) (Min et al. 2015) †
- ・etc.

† (参考) 長谷川忍: 「2010年代の学習支援研究におけるモデルのトレンド」, 人工知能, 35 (2), pp.250-256, 2020.

図 6. 「教育データサイエンスの可能性」の説明スライド例

2.4. 対象者

本研究の調査対象者は、セミナー参加者のうち第1回と第2回セミナー終了直後に行われた Web 調査に回答が得られた者であった (有効回答数: 第1回セミナー 48 名, 第2回セミナー 36 名)。

2.5. 調査内容

本セミナーの成果と課題を明らかにするために、(1) セミナーの満足度, (2) セミナー満足度の回答理由, (3) セミナーの難易度, (4) データサイエンスに関する知識習得の重要性, (5) 学校現場において教育データサイエンスを学習・活用することのメリット, (6) より知りたいセミナーの内容を調査内容とした。(1) は、質問項目「本日のセミナーは、どの程度満足いただけましたでしょうか」への回答を 5 件法 (1. 不満～5. 満足) にて求めた。(2) は、(1) で選択した満足度の理由を自由記述で求めた。(3) は、質問項目「今回のセミナーの難易度はどのように感じましたか」への回答を 5 件法 (1. とても易しかった～5. とても難しかった) にて求めた。(4) は、質問項目「今後の学校教育では、教員がデータサイエンスの知識を身に付けることは大切だと感じましたか」への回答を 5 件法 (1. まったく感じなかった～5. とても感じた) にて求めた。(5) は、質問項目「学校でデータサイエンスを活用することで、どんなことがわかる、できるようになると良いと思いますか?」にて該当する項目を全て選択するように求めた (具体的な項目は表 6 参照)。(6) は、セミナーの内容の中でより詳しく知りたいと感じたことを自由記述で求めた。なお、第1回セミナーでは (1), (2), (5) の一部、第2回セミナーでは (1) から (6) の回答を求めた。

2.6. 調査手続きと倫理的配慮

調査は第1, 2回それぞれのセミナー終了直後に Web フォームにて実施した。Web フォームの最初のページには、(1) 調査は無記名であること, (2) 調査結果は統計的に処理されるため個人が特定されることはないことを倫理的配慮として明記した。

2.7. 分析方法

第一に、セミナーの満足度と難易度、データサイエンスに関する知識習得の重要性について、得点の平均値をベイズ法により推定した。具体的には、長さ 21000 のチェーンを 5 つ発生させ、バーンイン期間を 1000 とし、ハミルトニアンモンテカルロ法によって得られた 100000 個の乱数で事後分布を近似した。事前分布として、母平均は一様分布 $U(1, 5)$, 母標準偏差は一様分布 $U(0, 5)$ に設定した。なお、ベイズ法による推定には `rstan(2.21.5)` を用いた。第二に、学校現場において教育データサイエンスを学習・活用することのメリットの特徴を検討するために、それぞれの項目を選択した人数と選択率を求めた。第三に、セミナー満足度の回答理由、より知りたいセミナーの内容の特徴を検討するために、これらに関する自由記述を整理した。

3. 「教育データサイエンスセミナー」の評価

3.1. セミナーの満足度と難易度、データサイエンスに関する知識習得の重要性の得点に関する平均値の推定

母数の全てについて、有効標本数は 44998 以上、収束判定指標は 1.00 以下であり、事後分布からの乱数の近似と考えられる。そこで、それぞれの母平均の事後期待値 (EAP), 事後標準偏差 (post.sd), 95% 信用区間 (95%CI) を表 5 に記す。表 5 から、以下 3 つの知見が得られた。

その 1 に、セミナーの満足度についてである。第1回セミナーでは母平均の事後期待値が 4.17, 95% 信用区間が [3.95, 4.39], 第2回セミナーでは事後期待値が 4.25, 95% 信用区間が [4.00, 4.50] といずれも 5 件法の意味的中央値 3.00 を上回る値であった。さらに、母平均の事後期待値が 3.00 より大きい確率は第1, 2回セミナーとも 100%, 4.00 より大きい確率は第1回セミナーが 93%, 第2回セミナーが 97% であった。以上から、第1, 2回セミナーとも参加者の満足度は高い傾向にあることが示された。よって、満足度の観点から、第1, 2回のセミナーの内容は教員養成・研修において適切なものと考えられる。

その 2 に、セミナーの難易度についてである。母平均の事後期待値が 3.64, 95% 信用区間が [3.40, 3.87] と 5 件法の意味的中央値 3.00 を上回る値であった。さらに、母平均の事後期待値が 3.00 より大きい確率は 100%, 4.00 より大きい確率は 0% であった。以上から、参加者は第2回セミナーの内容に対して、相対的に難しさを感じているが、その水準は高くないことが示された。よって、難しさの観点において、第2回セミナーの内容は教員養成・研修において適切なものと考えられる。

その 3 に、データサイエンスに関する知識習得の重要性についてである。母平均の事後期待値が 4.67, 95% 信用区間が [4.43, 4.89] と「4. まあまあ感じる」を上回る値であった。さらに、母平均の事後期待値が 4.00 より大きい確率は 100% であった。以上から、参加者は今後の学校教育においてデータサイエンスに関する知識

表 5. 平均値の事後分布に関する推定結果

	EAP	post.sd	95%CI	
			Lower	Upper
第1回セミナーの満足度	4.17	0.11	3.95	4.39
第2回セミナーの満足度	4.25	0.13	4.00	4.50
セミナーの難易度	3.64	0.12	3.40	3.87
データサイエンスに関する知識習得の重要性	4.67	0.12	4.43	4.89

表 6. 学校現場において教育データサイエンスを学習・活用することのメリット

	第1回 セミナー		第2回 セミナー	
	選択 人数	選択 率	選択 人数	選択 率
社会における AI・データサイエンス活用の動向がわかること	—	—	12	33%
基礎的な統計解析の理論や手法がわかること	—	—	21	58%
基礎的な AI、機械学習の仕組みがわかること	—	—	10	28%
教育における様々なデータをグラフなどで適切に可視化すること	—	—	19	53%
教育における様々なデータの代表値を求めるなど、基礎的な集計をすること	—	—	11	31%
全国規模の調査やテストの結果を自分の学校と比較すること	22	46%	14	39%
授業改善の効果を検証すること	43	90%	29	81%
児童生徒の実態を把握するために実施する様々なアンケートの結果を分析すること	31	65%	23	64%
保護者対象の様々なアンケートの結果を分析すること	21	44%	14	39%
自作したテスト問題やアンケートの適切性を検証すること	20	42%	16	44%
アンケートにおいて適切な調査票を作れること	—	—	14	39%
学習履歴から個別指導の必要な児童生徒を把握すること	27	56%	13	36%
AI 活用型ドリルアプリなど EdTech を活用すること	18	38%	8	22%
教育における新しい AI 活用のアイディアが考えられること	—	—	11	31%
教育における AI・データサイエンスの利用上の留意点がわかること	—	—	12	33%

を有する重要性を認識している傾向にあることが示された。よって、第2回セミナーの内容は今後の学校教育においてデータサイエンスに関する知識を習得することの重要性を認識させるものであったと推察される。

3.2. 学校現場において教育データサイエンスを学習・活用することのメリット

学校現場において教育データサイエンスを学習・活用することのメリットとして、それぞれの項目の選択人数と選択率を表6に記す。「授業改善の効果を検証すること」と「児童生徒の実態を把握するために実施する様々なアンケートの結果を分析すること」は、第1、2回両方のセミナーにおいて選択率が60%以上と相対的に高い水準であった。この結果から、第1、2回セミナーの内容は、授業改善の効果検証と児童生徒の実態把握において、教育データサイエンスを学習・活用することにメリットがあると認識させるものであったと考えられる。また、第2回セミナーのみで測定された「基礎的な統計解析の理論や手法がわかること」と「教育における様々なデータをグラフなどで適切に可視化すること」は、選択率が50%以上と相対的に高い水準であった。この結果は、第2回セミナーの内容は、統計的なデータ分析の理解とグラフなどによるデータの可視化において、教育データサイエンスを学習・活用することにメリットがあると認識させるものであったと考えられる。他方で、

「基礎的な AI、機械学習の仕組みがわかること」や「AI 活用型ドリルアプリなど EdTech を活用すること」など、AI に関する項目の選択率は相対的に低い傾向にあった。それゆえ、構想した「AI・データサイエンス基礎」と「教育データサイエンス」において、AI に関する知識の習得、ならびに教育における AI の活用について、より重きを置く必要があるだろう。

3.3. セミナー満足度の回答理由

まず、第1回セミナーの満足度の回答理由の自由記述として、代表的なものを表7に記す。表7より、基礎的な内容の習得・理解、セミナーのわかりやすさ、学校現場における教育データサイエンスの今後の展望に関する記述が得られた。

次に、第2回セミナーの満足度の回答理由の自由記述として、代表的なものを表8に記す。表8より、第1回セミナーと同様に基礎的な内容の習得・理解、セミナーのわかりやすさ、学校現場における教育データサイエンスの今後の展望に関する記述が得られた。さらに、第2回セミナーでは、内容の難しさや時間的制約に関する記述が得られた。

3.1と3.3の結果を統合すると、第1、2回ともセミナー満足度が高かったのは、参加者が教育データサイエンスに関する基礎的な内容と今度の展望を習得・理解できたこと、ならびにセミナーの内容が分かりやすかったこと

表 7. 第 1 回セミナーの満足度の回答理由

基礎的な内容の習得・理解	● 入門レベルの内容を知ることができた。
セミナーのわかりやすさ	● 基本的なことから学び直すことができた。
	● データサイエンスとは何か、AI とはなにか、記述統計の基礎など、データサイエンスの概要を理解するための必要な知識をわかりやすく解説していただき、大変参考になりました。
	● 難しい内容をできるだけ分かり易く、具体例を用いながら説明してくださったので。
今後の展望	● 今後、関連した情報に接する際に役立つと感じました。
	● 今後の教育界の変化について見通すことができた研修時間となりました。

(記述は原文のまま掲載)

表 8. 第 2 回セミナーの満足度の回答理由

基礎的な内容の習得・理解	● 推測統計（推定，検定），相関について知ることができた。
セミナーのわかりやすさ	● わからなかった抽象的なことが，具体的に理解することができた。
	● 素人にもわかりやすく説明していただけたから。
	● 統計についての知識・理解が乏しいが，図表や例を用いてがわかりやすく説明していただけたため。
今後の展望	● 現職教員にこれから必要になる情報と思われた。
	● 今後必要になる内容であるから。
内容の難しさ	● やや難解な部分もあったので。
	● 自分自身の勉強不足ですが，やや内容が難しく感じられました。
時間的制約	● 盛だくさんであったので，もう少し回数が多くてもよかったかなと感じています。
	● 少し内容が多く，難解な箇所がありましたが，このような試みは未来の学校教育において大変意義があると思います。

(記述は原文のまま掲載)

表 9. より知りたいセミナーの内容

データ収集の方法	● 学習データの取得方法。生徒の視線や動きなどを測るにはどうすれば良いか。
教育における具体的な事例	● 学校評価や学校でのアンケートの処理の具体
	● 学習ログをどのように活用していくのか具体例
データ分析の方法	● エクセルなどのデータ分析の事例をさらに取り入れていただきたい。
	● どのケースで，どの分析手法を使えばいいのか，実際には迷うことがあります。そのあたりの具体的なお話をいただくと助かります。

(記述は原文のまま掲載)

が要因であると考えられる。さらに，参加者は第 2 回セミナーの内容に難しさを感じていたが，第 1，2 回セミナーの満足度に顕著な差異が認められなかったこと（表 5）は興味深い結果である。この結果から，第 2 回セミナーで扱われた「統計的なデータ分析」，「教育データの読み取り方」，「教育データサイエンスの可能性」は学校教員にとって難しい内容であるものの，学校現場における重要性がとりわけ高いと認識した可能性が考えられる。それゆえ，「AI・データサイエンス基礎」と「教育データサイエンス」において，これらの内容はとりわけ重点的に指導することが求められよう。

3.4. より知りたいセミナーの内容

より知りたいセミナーの内容の自由記述として，代表的なものを表 9 に記す。表 9 より，データ収集の方法，教育における具体的な事例，データ分析の方法に関する記述が得られた。表 9 より，「AI・データサイエンス基礎」と「教育データサイエンス」の実施に向けて，以下 3 点の示唆が得られた。

その 1 に，データ分析だけではなく，データ収集の方法についても内容として組み込むことである。本セミナー，さらには構想した「AI・データサイエンス基礎」と「教育データサイエンス」において，データ収集に関する内容はほとんど扱われていない。そもそも，デー

タサイエンスにおいて，データ収集は主たる要素であるに関わらず，近年のデータサイエンスはデータ分析手法に過度に傾倒していることが問題であると指摘されている（松本，2022）。以上を踏まえると，データ分析だけではなくデータ収集の方法についても，「AI・データサイエンス基礎」と「教育データサイエンス」により組み込むことが求められよう。

その 2 に，教育における具体的な事例を多く組み込むことである。とりわけ，「教育データサイエンス」の第 2 回から第 8 回の内容は数理的側面の強いものであるため，教育における具体的な事例を適宜組み込むことが求められよう。

その 3 に，具体的なデータ分析の機会を多く設定することである。本セミナーでは，母平均の区間推定と検定においてのみ Excel での演習が行われたが，参加者はより多くの分析方法を習得することにニーズがあったものと考えられる。「教育データサイエンス」ではデータ分析の機会を多く設定されていることを踏まえると，「AI・データサイエンス基礎」において Excel などを用いたデータ分析の機会を多くすることが求められよう。

4. まとめ

兵庫教育大学の教育データサイエンスチームでは，教

員養成における AI・データサイエンス科目として学部学生全員が履修する「AI・データサイエンス基礎」と「教育データサイエンス」を構想した。その上で、この2科目の試行的実践として、統計解析や機械学習等に関する学習経験のない現職の学校教員を対象に「教育データサイエンスセミナー」を実施した。その結果、セミナーの満足度は高く、かつ難易度は適切な水準であることが示された。さらに、セミナーの内容は、データサイエンスに関する知識習得の重要性、ならびに教育データサイエンスを学習・活用することにメリットがあることを認識させるものであった。以上から、構想した「AI・データサイエンス基礎」と「教育データサイエンス」の内容は、教員養成における AI・データサイエンス科目として適切なものと考えられる。ただし、構想した内容の改善点として、データ収集の手法を内容として取り入れること、AIに関する知識の習得と教育における AI の活用に重きを置くことが示された。これらの改善点に即して、「AI・データサイエンス基礎」と「教育データサイエンス」の内容を修正・実施・評価することが今後の課題となる。

文 献

- Graduate School of Education, Stanford University (n.d.)
Education Data Science Program (Master's Program) .
<https://ed.stanford.edu/eds/program> (参照日 2022.07.12)
- 国立教育政策研究所 (n.d.) 教育データサイエンスセンター https://www.nier.go.jp/04_kenkyu_annai/div12-data-sci.html (参照日 2022.07.12)
- 松本渉 (2022) データサイエンスの忘れ物. 情報研究：関西大学総合情報学部紀要, 54, 81-93.
- 文部科学省 (2017a) 小学校学習指導要領 (平成 29 年告示) 解説 総則編. https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/fieldfile/2019/03/18/1387017_001.pdf (参照日 2022.07.08)
- 文部科学省 (2017b) 小学校学習指導要領 (平成 29 年告示) 解説 算数編. https://www.mext.go.jp/content/20211102-mxt_kyoiku02-100002607_04.pdf (参照日 2022.07.12)
- 文部科学省 (2018a) 高等学校学習指導要領 (平成 30 年告示) 解説 情報編. https://www.mext.go.jp/content/1407073_11_1_2.pdf (参照日 2022.07.12)
- 文部科学省 (2018b) 高等学校学習指導要領 (平成 30 年告示) 解説 数学編 理数編. https://www.mext.go.jp/content/1407073_05_1_2.pdf (参照日 2022.07.12)
- 文部科学省 (2019) 新時代の学びを支える先端技術活用推進方策 (最終まとめ). https://www.mext.go.jp/component/a_menu/other/detail/__icsFiles/fieldfile/2019/06/24/1418387_02.pdf (参照日 2022.07.12)
- 内閣府 (2019) AI 戦略 2019：人・産業・地域・政府全てに AI. <https://www8.cao.go.jp/cstp/ai/aistrategy2019.pdf> (参照日 2022.07.12)
- 数理・データサイエンス教育強化拠点コンソーシアム

(2020) 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.
http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf (参照日 2022.07.12)