

## 高大接続を見据えた「情報Ⅰ」からみるデータサイエンス教育の関係分析 Relationship Analysis of Data Science Education from the perspective of Informatics I for High School-University Connection

林 宏樹<sup>1</sup> 井手 広康<sup>2</sup> 渡辺 博芳<sup>3</sup>  
Hiroki Hayashi Hiroyasu Ide Hiroyoshi Watanabe

### 1. はじめに

#### 1.1 背景

現代社会において、数理・データサイエンス（以下、DS）・AIの素養は、文理を問わず全ての学修者にとって不可欠なものとなりつつある。この潮流を受け、大学・高等専門学校教育の指針として、数理・データサイエンス・AI教育強化拠点コンソーシアム（以下、コンソーシアム）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」（以下、モデルカリキュラム）を策定・公開している[1]。同カリキュラムは、学生がデジタル社会においてDS・AIを基礎的素養として主体的に活用できることを目指しており、特に「コア学修項目」として、データを適切に読み解き（2-1）、説明し（2-2）、扱う（2-3）能力、すなわち基本的なDSに関する力の育成を必須としている。

一方、高等学校においては、2022年度より新学習指導要領（平成30年告示）（以下、学習指導要領）が施行され、「情報Ⅰ」が共通必修科目となった[2]。同科目においても、「(4)情報通信ネットワークとデータの活用」が主要な学習内容の1つであり、データに基づいた科学的な思考力や判断力を養うことが目標とされている。2025年度からは、この「情報Ⅰ」が大学入学共通テスト（以下、共通テスト）の出題科目となった。一般に、共通テストのような大規模選抜試験は、高等学校の教育課程や授業実践に対して「ウォッシュバック効果（波及効果）」と呼ばれる影響を与えることが指摘されている。共通テストが学習指導要領に準拠する以上、その出題内容は高校現場における指導の重点化を促す可能性がある。したがって、共通テスト「情報Ⅰ」の導入は、高等学校におけるデータ分析能力育成の方向性に影響を与え、ひいては大学でのDS・AI教育への接続にも関わる重要な要素となり得る。

#### 1.2 研究目的と意義

高等学校「情報Ⅰ」と大学におけるDS・AI教育（リテラシーレベル）との円滑な接続は、体系的なDS人材育成において極めて重要である。本研究では、この高大接続の現状と課題を明らかにすることを目的とし、以下の二点を分析する。第一に、高等学校学習指導要領解説情報編[2]における「情報Ⅰ」の「データの活用」に関する内容と、コンソーシアムのモデルカリキュラム（リテラシーレベル）の内容を比較検討し、両者の対応関係、接続性およびギャップを特定する。第二に、共通テスト「情報Ⅰ」に関して

公開されている問題セットの「データの活用」分野を分析対象とし、モデルカリキュラムのデータ分析に関する学修項目との対応関係を評価する。これらの分析を通じて、共通テスト「情報Ⅰ」が測定しようとしているDSに関する範囲を見出すとともに、高等学校「情報Ⅰ」から大学におけるDS・AI教育への接続性を評価する。本研究の成果は、今後のDS教育カリキュラムの開発・改善、および高大接続の強化に資する基礎的知見を提供することが期待される。

### 2. 先行研究

わが国では、大学教育を含む高等教育改革の一環として「高大接続改革」[3]が提言され、高等学校教育・大学教育・大学入学選抜を一体的に見直す必要性が指摘されてきた。中央教育審議会[4]は同改革を「教育改革における最大の課題でありながら実現が困難」であると位置づけ、その実現がわが国の人材育成に不可欠であると強調している。

情報分野では、小・中・高を縦断した体系的カリキュラムの構築が求められているが、高等学校「情報」と大学初年次の情報基礎教育との連続性の乏しさがかねてより指摘されてきた。村松[5]は、この課題を解消するには縦（小中高大）の接続と横（学会・教員養成）の連携を同時に進める必要があると論じる。実際、高校生の情報活用能力と大学初年次生のそれを比較した実証研究[6]でも、能力ギャップの存在が報告されている。

共通テストは、高大接続改革の一環として従来の大学入試センター試験に代わり導入された制度である。共通テストは「大学教育を受けるために必要な基礎的・汎用的能力を把握する」ことを目的とし、全国の大学が共同で実施する選抜手段として位置づけられている[7]。なかでも「情報」の出題が加わった背景には、国立大学協会が掲げる「6教科8科目」方針と、それを後押しした文部科学省の報告書「大学における数理及びデータサイエンス教育の方策」（2016年12月）が大きく影響したと指摘されている[8]。同時期、文部科学省は6つの大学を拠点校に選定してコンソーシアムを形成し、当該組織がモデルカリキュラムを策定した[1]。さらに、清水[9]は、学習指導要領「情報Ⅰ」「情報Ⅱ」に含まれるDS関連内容が、このモデルカリキュラムのリテラシーレベルと整合的であると論じている。よって、共通テストは高大接続を具現化する1つの鍵と考えることができる。したがって、DS教育の高大接続に関する1つの指標として、共通テストとモデルカリキュラムの対応関係を分析する意義はあると考えられる。

「情報Ⅰ」とDS教育の接続に関する先行研究としては、学習指導要領・教科書をもとに共通教科情報と大学の情報教育の接続を検討した事例[10]や、「情報Ⅰ」教科書におけるDS項目を体系的に考察した研究[11]が報告されている。しかし、これらは共通テストとの関連性を直接扱っていない。

1 雲雀丘学園中学校・高等学校  
Hibarigaoka Gakuen Junior & High School  
2 愛知県立旭丘高等学校  
Aichi Prefectural Asahigaoka High School  
3 電気通信大学  
The University of Electro-Communications

また、井手[12]はプログラミング分野に限定し、教科書内容と大学入試センターが公表した試作問題・サンプル問題との整合を分析しているものの、データサイエンス分野には言及していない。以上を踏まえ、本研究では情報Ⅰの「データの活用」分野に焦点を当て、共通テストの関連設問と DS リテラシー・モデルカリキュラムとの対応関係を実証的に検証する。

### 3. 分析の枠組み

#### 3.1 モデルカリキュラム（リテラシーレベル）

本研究で参照するモデルカリキュラム[1]は、数理・データサイエンス・AI 教育強化拠点コンソーシアムにより策定され、大学・高等専門学校生が共通して身につけるべき DS・AI の基礎的素養を示したものである。その中核をなす「コア学修項目」には、「1.社会におけるデータ・AI 利活用」「2.データリテラシー」「3.データ・AI 利活用における留意事項」が含まれる。本研究では、特に「2.データリテラシー」に着目する。これは、「2-1.データを読む」「2-2.データを説明する」「2-3.データを扱う」の3つのサブ項目から構成され、データの特徴の読解、他者への適切な説明、基礎的な集計・加工能力の育成を目的としている。さらに、オプションとして「4.スキルセット」が設定されており、例えば「4-1.統計および数理基礎」「4-4.時系列データ解析」「4-7.データハンドリング」「4-8.データ活用実践（教師あり学習）」など、より発展的な内容も含まれる。本研究の分析に特に関連するデータリテラシーを付録 A、オプション項目を付録 B に示す。

#### 3.2 学習指導要領「情報Ⅰ」における「データの活用」

高等学校学習指導要領解説[2]によれば、「情報Ⅰ」の「(4)情報通信ネットワークとデータの活用」における「データの活用」では、データに基づいた問題発見・解決能力の育成を目指している。具体的に、学習指導要領解説[2] (pp.35p-40) の「データの活用」分野と判断できるア（ウ）（A と表記）、イ（ウ）（B と表記）に記載されている 1 文 1 文の内容に基づき、以下に学習内容をまとめる。なお、学習指導要領の 1 文ずつ要約し、A の観点における 1 行目を A1 と記載する。また、全体にわたる学習活動 (pp.40) において、A、B と重複していない内容（C と表記）についても同様にまとめる。

##### A 身に付ける知識及び技能 (pp. 38)

- A1 データの様々な形式、データの収集、整理、分析する一連のデータ処理の流れ及び評価に関する理解など。
- A2 様々な形式のデータを扱うなど。
- A3 尺度水準（名義・順序・間隔・比例）の理解、質的データ、量的データの扱いの違いなど。
- A4 データの内容や形式に応じたデータ収集方法など。
- A5 データの整理として欠損値や外れ値の整理など。
- A6 データ分析として、可視化の方法、テキストマイニングなどを含む基礎的な分析可視化の方法

##### B 身に付ける思考力、判断力、表現力 (pp. 38-40)

- B1 必要なデータ収集に関する選択、判断する力、適切なデータにおける整理、変換、分析、処理、可視化の方法を選択、判断する力など。
- B2 データの傾向を評価し、客観的な指標を基に判断する力、適性な解釈を行う力など。

表 1 学習指導要領とモデルカリキュラムの対応関係

| 要領 | 項目  | キーワード                                                      |
|----|-----|------------------------------------------------------------|
| A1 | 2-1 | 全体                                                         |
| B1 | 4-8 | 全体                                                         |
| A2 | 2-1 | データの種類、データの分布と代表値、代表値の性質の違い                                |
| A4 | 2-1 | データのばらつき、外れ値、観測データに含まれる誤差、打ち切りや欠測を含むデータ、層別の必要なデータ、母集団と標本抽出 |
|    | 2-3 | データの取得                                                     |
|    | 4-8 | データの収集                                                     |
| A5 | 2-1 | データのばらつき、外れ値                                               |
|    | 4-8 | データの加工                                                     |
| A6 | 2-2 | データ表現、不適切なグラフ表現                                            |
| B2 | 2-1 | 統計情報の正しい理解                                                 |
|    | 4-8 | データ分析結果の共有                                                 |
| B3 | 2-1 | 外れ値、母集団と標本抽出                                               |
|    | 4-8 | データの収集                                                     |
| B4 | 4-4 | 時系列データ、季節調整、移動平均                                           |
| B5 | 2-1 | データのばらつき、データ母集団と標本抽出                                       |
|    | 4-8 | データの加工                                                     |
| B6 |     | 該当なし                                                       |
| C1 | 2-1 | 相関と因果                                                      |
|    | 4-8 | データの分析                                                     |
| C2 | 4-8 | データの加工                                                     |
| C3 | 2-1 | 散布図行列                                                      |
|    | 2-3 | データ解析ツール                                                   |
|    | 4-8 | データの分析                                                     |

B3 データ収集前の分析の構想の練り紐づける項目の洗い出し、外れ値の扱い、仮説検定の考え方等を扱うなど。

B4 時系列データ、個人と個人の繋がりを表現するデータ等を扱うなど。

B5 オープンデータの扱い、データ収集の偏り、ソフトウェアを扱った整理、加工、分析、可視化など。

B6 単語の重要度や関係性等のテキストマイニングなど。

##### C 全体にわたる学習活動 (pp. 40)

C1 相関係数などの統計指標、相関関係や因果関係、交絡因子、単回帰分析など。

C2 データに必要な事前処理、データの傾向などを読む工夫など。

C3 分析・可視化のためのソフトウェアの活用、複数の散布図、相関関係の見られる変数の組み合わせ、回帰直線によるデータの変化の予測など。

#### 3.3 学習指導要領とモデルカリキュラムの対応

3.2 節に基づき、モデルカリキュラムの学修項目との対応関係をみる。そこで、付録 A「2.データリテラシー」のキーワードとの対応関係を表 1 に示す。なお、表 1 では 3.2 節の学習指導要領を要領と表記し、付録 A の項目とキーワードとの対応関係を示した。なお、モデルカリキュラムの「データ活用実践（教師なし学習）」では「4-8.データの活用実践（教師あり学習）」と重複するキーワードがある。

が、教師なし学習の内容は「情報Ⅰ」の内容には含まれていないことを考慮した。結果、学習指導要領とモデルカリキュラムの対応関係は2-1,2-2,2-3,4-8にあると判断した。

### 3.4 高等学校「情報Ⅰ」と大学DS・AI教育の接続

上記の分析から、高等学校「情報Ⅰ」における「データの活用」分野は、大学におけるDS・AI教育（リテラシーレベル）の基礎を形成する上で重要な役割を担っていることが確認できる。特に、データの種類や統計量の基本的な理解、多様なグラフを読み解き作成する能力、相関の概念把握などは、モデルカリキュラムのコア項目である「2-1.データを読む」「2-2.データを説明する」能力の基盤となる。高等学校教育を通じて一定のデータリテラシーが育成されることが期待される。

一方で、いくつかのギャップも存在する。第一に、分析手法の深さである。相関と因果の区別や交絡因子の考慮、回帰分析の応用などは、高等学校では限定的な扱いとなる可能性が高いが、大学リテラシーレベルではより明確に意識されるべき項目である。ここで、データの関係性については相関関係が取り上げられ、相関係数で関係性を判断できることが示されている。散布図との組み合わせで、正の相関、負の相関、相関なしというデータの関係性の程度について説明されている。相関係数の計算について取り扱っている教科書は、7種類である[11]。また散布図との関連で6種類の教科書で回帰分析についての説明がある。相関関係や因果関係との関連で回帰直線について触れている。第二に、扱うデータの種類と量、そしてデータハンドリングのスキルである。高等学校「情報Ⅰ」では主に表形式データを対象とし、表計算ソフトでの扱いが中心となるが、大学では非構造化データやより大規模なデータを扱い、データベース（SQL）やプログラミング言語（Python, R など）を用いたデータ処理・加工スキル（モデルカリキュラムオプション項目4-7など）の習得が求められる場面が増える。第三に、教育目標の重点である。モデルカリキュラムは「データ思考の涵養」を掲げ、社会におけるデータ・AI利活用の文脈や、それに伴う倫理的・法的・社会的課題（ELSI）への意識をより強く求めている点が、高等学校の学習指導要領解説の記述と比較して特徴的である。

これらの分析は、高等学校「情報Ⅰ」で育成される力が、大学でのデータサイエンス学習の重要な土台となる一方で、大学教育においては、より高度な分析手法、実践的なデータハンドリングスキル、そしてデータ利活用に伴うELSIへの深い理解などを接続的に、かつ発展的に育成していく必要があることを示している。

### 3.5 共通テストとの関連に対する示唆

高等学校「情報Ⅰ」の「データの活用」分野に関する学習内容は、表1の結果よりモデルカリキュラムの2-1,2-2,2-3,4-4,4-8で示せることがわかった。特に「2-1.データを読む」「2-2.データを説明する」能力と強い関連性を有しており、高大接続の観点からその重要性は高い。高校生は「情報Ⅰ」を通じて、データ分析の基本的な概念とスキルを習得し、大学でのより専門的な学習への準備を行うことができる。

一方、「2-3.データを扱う」は技能の要素が大きく、共通テストはペーパーテストであることを考慮し、測定することが困難であると判断した。

以上のことから、モデルカリキュラムと共通テストの関係については、「2-1.データを読む」「2-2.データを説明する」「4.4.時系列データ解析」「4-8.データ活用実践（教師あり学習）」との対応関係をみることにした。

## 4. 共通テスト「情報Ⅰ」問題の分析

### 4.1 分析対象と方法

本章では、前章で確認した高等学校「情報Ⅰ」とモデルカリキュラムの関係性を踏まえ、実際の共通テスト「情報Ⅰ」の出題がモデルカリキュラムとどのように対応するかを分析する。分析対象は、第1章で述べた共通テスト「情報Ⅰ」に関連する次の(1)～(6)に示す公開問題6種類（検討用イメージ、サンプル問題、試作問題、参考問題、本試験、追・再試験）における「データの活用」分野（主に第3問または第4問）である。

- (1) 「情報」試作問題（検討用イメージ）（以下、検討用イメージ）[13]
- (2) 平成30年告示高等学校学習指導要領に対応した令和7年度大学入学共通テスト「情報」サンプル問題（以下、サンプル問題）[14]
- (3) 令和7年度大学入学共通テスト試作問題「情報Ⅰ」（以下、試作問題）[15]
- (4) 令和7年度大学入学共通テスト「情報Ⅰ」試作問題（参考問題）（以下、参考問題）[16]
- (5) 令和7年度大学入学共通テスト「情報Ⅰ」本試験問題（以下、本試験）[17]
- (6) 令和7年度大学入学共通テスト「情報Ⅰ」追・再試験問題（以下、追・再試験）[18]

分析方法として、上記6種類の問題セットについて、各設問の主題、題材、問われている内容等を精査した。その上で、各設問が測定しようとしている知識・スキルを、付録A、付録Bに示したモデルカリキュラムのキーワードと照合し、対応関係を評価した。

評価方法は、3人の著者が独立して評価を行い、集計し、一致しなかった項目があった場合は協議を行い決定した。

対応したキーワードは、付録A、付録Bのキーワードの前に記載した2-1-1のような表記で示す。1つの設問が複数の学修項目に関連する場合は、すべて記録した。分析の客観性を担保するため、各問題セットの概要および設問ごとの分析結果を基礎資料として用いた[19][20]。

### 4.2 分析結果

#### (1) 検討用イメージ（第8問C）

SNSの発信データを題材として、データの収集・整理・分析に関する基本的な知識・技能と、データやグラフから読み取れることを考察する力を問う構成となっている。SNSの情報発信件数とWebサイト訪問回数の月別データ（表や散布図）を提示し、これらの関係性（相関の有無や傾向）を読み解く力を問うている。さらに、示された回帰直線とその式から、回帰係数の意味を理解し、具体的な数値を解釈できるかを問うている。よって、2-1-5（相関と因果）、2-2-1（データ表現）、4-8-4（データの分析）が該当すると判断した。

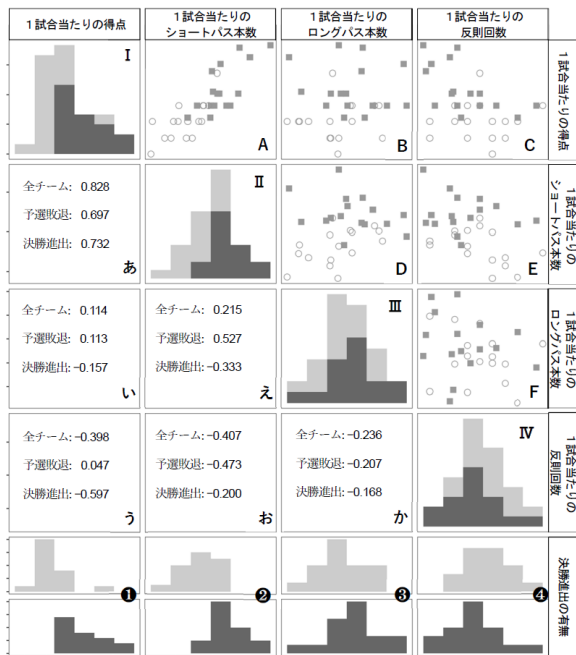


図1 サンプル問題に出題された散布図行列[14]

## (2) サンプル問題（第3問）

オープンデータであるサッカーのワールドカップ出場チームのデータ（試合数、総得点、パス本数、反則回数など）を用いた問題解決活動を通じて、データの活用に関する考察力を問うている。具体的には、予選敗退チームと決勝進出チームのグループに分け、表計算ソフトウェア等でデータを整理・加工（1試合あたりの数値に変換）した上で、統計処理ソフトウェアで作成された図1に示す散布図行列（散布図、相関係数、ヒストグラムを含む）や基本統計量の表を用いて、項目間の関係性やグループ間の傾向の違いを読み取る能力を評価している。さらに、回帰直線を用いた予測値や残差の計算、クロス集計表の解釈などを通じて、実践的なデータ分析と考察力を問うている。

問1：4つの変数（1試合あたりの得点、ショートパス本数、ロングパス本数、反則回数）について、グループ別（決勝進出・予選敗退）に色分け・マーカー分けされた散布図、グループ別および全体の相関係数、グループ別のヒストグラムを組み合わせた散布図行列が提示されている。この複雑な図1から、特定のグループにおける変数間の相関（例：決勝進出チームにおける負の相関）、グループ間で相関の傾向が異なる組み合わせ、各図（散布図、ヒストグラム）の表現内容などを正確に読み取り、解釈する力を問うている。よって、2-1-2（データの分布と代表値）、2-1-5（相関と因果）、2-1-9（散布図行列）、2-2-1（データの表現）、決勝チームと予選敗退チームによる層別の判断する選択肢から2-1-7（層別の必要なデータ）と判断した。

問2：問1で着目した「1試合あたりの得点」と「1試合あたりのショートパス本数」の関係について、決勝進出チームと予選敗退チームそれぞれについて回帰直線を示した散布図が提示されている。これらの回帰直線の傾きから得点の増加率の違いを比較したり、特定の値（ショートパス本数）における予測値の差を計算したり、実際の値と回帰直線による予測値との差（残差）を計算したりする力を問

うている。単回帰分析に基づくデータの解釈と予測に関する考察力を評価している。よって、2-1-7（層別の必要なデータ）、4-8-4（データの分析）と判断した。

問3：決勝進出チームと予選敗退チームそれぞれについて、4つの変数の基本統計量（最小値、四分位数、最大値、分散、標準偏差、平均値）をまとめた表が提示されている。この表から、四分位範囲や標準偏差といった散らばりの指標、中央値と平均値の関係などを読み取り、グループ間のデータの分布の特徴を比較・考察する力を問うている。よって、2-1-2（データの分布と代表値）、2-1-4（データのばらつき）に該当すると判断した。

問4：「決勝進出の有無」と「1試合あたりの反則回数」（全チームの四分位数で3つのカテゴリに分類）の関係を分析している。散布図行列や統計量表から読み取れる全体的な傾向を把握した上で、クロス集計表の空欄を埋め、その表から特定の条件における割合を計算するなど、カテゴリカルデータの分析と解釈に関する力を問うている。よって、2-1-2（データの分布と代表値）、2-1-5（相関と因果）、2-1-9（クロス集計、散布図行列）と判断した。

## (3) 試作問題（第4問）

国が実施した生活時間の実態に関する統計調査を基に、15歳以上19歳以下の若年層における都道府県別の平日1日の生活行動時間データを用いている。特に、スマートフォン・パソコンなどの使用時間が「1時間未満」のグループ（短いグループ）と「3時間以上6時間未満」のグループ（長いグループ）に分け、使用時間と睡眠・学業の時間との関係を題材としている。データの活用と分析に関する基本的な知識・技能、およびデータが表すグラフから読み取れることを考察する能力を問うている。

問1：短いグループと長いグループの統計データ（表1-A、表1-B）のみを基にして、分析可能な仮説とそうではない仮説（例：使用時間帯や具体的な行動内容など、データに含まれない情報に基づく仮説）を識別できるかを問うている。よって、4-8-2（データの収集）と判断した。

問2：両グループの「睡眠の時間」と「学業の時間」の分布をまとめた箱ひげ図から、中央値、四分位数、範囲、外れ値などのデータの分布の特徴を正確に読み取ることができると問うている。よって、箱ひげ図に基づいて、データの分布を把握する問いであり、2-1-2（データの分布と代表値）、2-2-1（データ表現）に該当すると判断した。

問3：都道府県ごとに「短いグループ」の値から「長いグループ」の値を引いた「睡眠の時間」および「学業の時間」の差を箱ひげ図で示している。この図から、スマートフォン・パソコンなどの使用時間の長さの違いが、睡眠時間と学業時間のどちらにより顕著な差（影響）をもたらしているかを考察できるかを問うている。よって、2-2-1（データ表現）、2-2-2（データの比較）に該当すると判断した。

問4：短いグループにおける「学業の時間」と「睡眠の時間」の関係を表した散布図を提示し、全体的な相関（ここでは弱い負の相関）の傾向について、その解釈の妥当性を問うている。よって、選択肢を考慮し、2-1-4（データのばらつき）、2-1-5（相関と因果）、2-2-1（データ表現）と判断した。

問5：問4の散布図に回帰直線を加えた図と、横軸に回帰直線による睡眠時間の推定値、縦軸に残差（実際の睡眠時間－推定値）を標準化した値をとったグラフを提示して

いる。これらの図から、回帰直線からの個々のデータのずれ（残差）の大きさを読み取り、標準偏差を基準とした外れ値の判定（ここでは標準偏差の 2 倍以上）を理解・適用できるかを問うている。よって、2-1-4（データのばらつき、外れ値）、2-2-1（データ表現）、4-8-3（データの加工）、4-8-4（データの分析）と判断した。

#### (4) 参考問題（第 4 問）

「情報 I」の授業におけるデータ分析の課題として、季節と商品の売り上げの関係性を調べるというテーマ設定である。具体的には、エアコンの全国月別売上台数と、ある K 市におけるアイスクリームの月別売上個数のデータ（2016 年 1 月～2020 年 12 月）を用いている。これらの時系列データや、追加された K 市の月別平均気温・平均湿度データを用いて、グラフの解釈、移動平均、変数間の相関（時間差を含む）、散布図行列などの統計的な見方や分析手法に関する理解を問うている。

問 1：エアコンとアイスクリームの月別売上数を重ねて表示した折れ線グラフから、期間内の傾向（増加・減少）、季節性（ピーク時期のずれ）、特定時点での値の比較などを正確に読み取れるかを問うている。よって、2-2-1（データ表現）、4-4-1（時系列データ）と判断した。

問 2：月ごとの変動をならして長期的な傾向を見るための「移動平均」の概念を説明し、異なる期間（6 か月、9 か月、12 か月、15 か月）で計算された移動平均のグラフの中から、12 か月移動平均に該当するグラフを特定できるかを問うている。季節変動の周期（12 か月）と移動平均期間の関係を理解しているかがポイントとなる。よって、2-2-1（データ表現）、4-4-2（移動平均）と判断した。

問 3：ある時系列データとそのデータを  $n$  か月ずらしたデータとの相関係数をグラフ（コレログラム）で示し、そのグラフからデータの周期性（エアコン売上の場合は約 12 か月周期）を読み取ることができるかを問うている。よって、2-1-5（相関と因果）、2-2-4（優れた可視化事例の紹介）と判断した。

問 4：エアコンとアイスクリームの売上数の相関関係について、時間的なずれ（ラグ）を考慮して分析している。エアコンのデータを 1 か月後にずらした場合に相関係数が最も高くなる（0.62→0.86）という結果から、両者の売上ピークの時期的なずれや、一方のデータを使った他方の予測可能性について考察できるかを問うている。よって、予測可能性を検討、つまり 4-8-4 のモデルの評価を行っており、相関関係と因果関係を含めた選択肢があるため、2-1-5（相関と因果）と 4-8-4（データの分析）と判断した。

問 5：エアコン売上、アイスクリーム売上、平均気温、平均湿度の 4 つの変数について、それぞれのヒストグラム、変数間の散布図、相関係数を一覧表示した散布図行列を提示している。この図から各変数間の相関の強弱を読み取り、それらの関係性をモデル化し、相関関係や潜在変数を読み取ることができるかを問うている。よって、2-1-5（相関と因果）、2-1-9（散布図行列）と判断した。

#### (5) 本試験（第 4 問）

観光庁が公開している 2019 年の「旅行・観光消費動向調査」のデータを用いて、旅行に関する実態を分析する場面設定である。地方別および都道府県別の延べ旅行者数を、旅行の目的（出張等、帰省等、観光等）別に分析している。データの尺度水準の理解、棒グラフや帯グラフ、散布図、

箱ひげ図といった様々なグラフの読解、相関の解釈、人口あたり指標のような派生データの活用、四分位数を用いたカテゴリ分類と比較など、データ分析に関する多角的な知識と考察力を問うている。

問 1：地方別の目的別旅行者数データと、それに基づく棒グラフ・帯グラフが提示されている。データの尺度水準（名義尺度、比例尺度）について正しく理解しているかを確認するとともに、グラフから地方ごとの旅行者数の絶対数や目的別割合の比較・読解ができるかを問うている。よって、2-1-1（データの種類）、2-2-1（データ表現）と判断した。

問 2：都道府県別の目的別旅行者数データと、目的間の組み合わせによる散布図および相関係数が提示されている。これらの散布図と相関係数から読み取れることを正しく解釈できるかを問うている。よって、2-1-5（相関と因果）、2-2-1（データ表現）と判断した。

問 3：都道府県別の「出張等」と「観光等」の旅行者数に着目し、その散布図と、各旅行者数を人口で割った指標「出張/人口」「観光/人口」の散布図を比較している。両方の散布図に  $y=x$  の直線を加えることで、どちらの目的の旅行者が相対的に多いかを視覚化し、人口で割ることによる指標の変化（特に特定の都道府県を示す点の相対的位置の変化）の理由を考察できるかを問うている。よって、2-2-1（データ表現）、2-2-2（データの比較）、4-8-3（データの加工）と判断した。

問 4：「出張/人口」と「観光/人口」の散布図に、各軸の分布を示す箱ひげ図を併記した図を使用している。特定の閾値を超える都道府県数を読み取ったり、第 3 四分位数などを基準に都道府県をカテゴリ分類し、そのカテゴリに該当する都道府県数を比較したり、特定のカテゴリに属する都道府県をグラフ上の点（A～F）から特定したりする力を問うている。よって、2-1-4（外れ値）、2-1-7（層別の必要なデータ）、2-2-1（データ表現）と判断した。

#### (6) 追・再試験（第 4 問）

情報の授業の研究課題として「高齢化」をテーマとし、Web 上のオープンデータから収集した 47 都道府県のデータを用いて、「高齢化と病院数との関係」を分析する場面設定である。データには、総人口、年齢 3 区分別人口（15 歳未満、15～64 歳、65 歳以上）、面積（可住地面積）、病院数が含まれている。これらの元データや、そこから計算される各種の指標（高齢化率、人口あたり・面積あたり病院数、人口密度など）を用いて、相関分析、散布図行列の解釈、層別分析、回帰分析といったデータ分析の手法や考え方についての理解を問うている。

問 1：3 つの年齢層別人口と病院数の 4 変数間の相関係数行列を示し、相関係数の基本的な性質（絶対値と分布の関係など）の理解と、相関行列から読み取れる全体的な傾向（人口が多い都道府県ほど病院数も多い）を把握できるかを問うている。よって、2-1-5（相関と因果）、2-1-9（相関係数行列）と判断した。

問 2：新たに「65 歳以上割合」「人口 1 万人あたり病院数（病院/人口）」「面積 1km<sup>2</sup>あたり病院数（病院/面積）」という 3 つの指標を導入し、それらの散布図行列（ヒストグラム、散布図、相関係数を含む）を提示している。「病院/面積」が示す意味を解釈させるとともに、「65 歳以上割合」と他の 2 指標との相関の符号の違いから、高齢化が

表2 6種類の問題セットとキーワードとの対応関係

|                               | (1)検 | (2)サ | (3)試 | (4)参 | (5)本 | (6)追 | 全体 |
|-------------------------------|------|------|------|------|------|------|----|
| 2-1-1 データの種類                  |      |      |      |      | ○    |      | ○  |
| 2-1-2 データの分布                  |      | ○    | ○    | ○    |      | ○    | ○  |
| 2-1-3 代表値の性質の違い               |      |      |      |      |      |      |    |
| 2-1-4 データのばらつき                |      | ○    | ○    |      | ○    |      | ○  |
| 2-1-5 相関と因果                   | ○    | ○    | ○    | ○    | ○    | ○    | ○  |
| 2-1-6 観測データに含まれる誤差の扱い         |      |      |      |      |      |      |    |
| 2-1-7 打ち切りや欠測を含むデータ，層別の必要なデータ |      | ○    | ○    |      | ○    | ○    | ○  |
| 2-1-8 母集団と標本抽出                |      |      |      |      |      |      |    |
| 2-1-9 クロス集計表，分割表，相関係数行列，散布図行列 |      | ○    |      | ○    |      | ○    | ○  |
| 2-1-10 統計情報の正しい理解             |      |      |      |      |      |      |    |
| 2-2-1 データ表現                   | ○    | ○    | ○    | ○    | ○    | ○    | ○  |
| 2-2-2 データの比較                  |      | ○    | ○    | ○    | ○    | ○    | ○  |
| 2-2-3 不適切なグラフ表現               |      |      |      |      |      |      |    |
| 2-2-4 優れた可視化事例の紹介             |      | ○    | ○    | ○    | ○    | ○    | ○  |
| 2-2-5 相手の的確かつ正確に情報を伝える技術や考え方  |      |      |      |      |      |      |    |
| 4-4-1 時系列データ                  |      |      |      | ○    |      |      | ○  |
| 4-4-2 季節調整，移動平均               |      |      |      | ○    |      |      | ○  |
| 4-8-1 教師あり学習による予測             |      |      |      |      |      |      |    |
| 4-8-2 データの収集                  | ○    | ○    | ○    | ○    | ○    | ○    | ○  |
| 4-8-3 データの加工                  |      |      | ○    | ○    | ○    | ○    | ○  |
| 4-8-4 データの分析                  | ○    | ○    | ○    | ○    |      | ○    | ○  |
| 4-8-5 データ分析結果の共有，課題解決に向けた提案   |      |      |      |      |      |      |    |

※検：検討用イメージ，サ：サンプル問題，試：試作問題，参：参考問題，本：本試験，追：追・再試験

進んでいる都道府県における人口あたり・面積あたり病院数の傾向を考察できるかを問うている。よって，2-1-5（相関と因果），2-1-9（散布図行列，相関係数行列），2-2-1（データ表現），2-2-2（データの比較）と判断した。

問3：さらに「人口密度」を指標に加え，4変数（人口密度，65歳以上割合，病院/人口，病院/面積）の散布図行列を提示している。この行列から読み取れる相関関係をもとに，人口の過密・過疎と高齢化，および病院数指標との関係性についての仮説を構築・選択できるかを問うている。よって，2-1-5（相関と因果），2-1-9（散布図行列，相関係数行列），2-2-1（データ表現），2-2-2（データの比較）と判断した。

問4：「人口密度」の中央値で都道府県を2グループ（X，Y）に分け，「65歳以上割合」と「病院/面積」の散布図（グループ別に表示）を提示している。「人口密度」のヒストグラムの形状を解釈させるとともに，散布図から2つのグループの分布の特徴（平均値，範囲，傾向の違いなど）を読み取ることができるかを問うている。よって，2-1-2（データの分布と代表値），2-1-7（層別の必要なデータ），2-2-1（データ表現）と判断した。

問5：「人口密度」から「病院/面積」を予測するための回帰直線を示し，与えられた人口密度の値を用いて，回帰式から病院/面積の予測値を計算できるかを問うている。よって，4-8-4（データの分析）と判断した。

以上，6種類の問題セットとキーワードとの対応関係の結果を表2にまとめた。

## 5. 考察

前章の分析結果に基づき，共通テスト「情報Ⅰ」の「データの活用」分野とモデルカリキュラム（リテラシーレベル）との関連性について考察する。

### 5.1 モデルカリキュラムからみる共通テストの内容

分析対象とした6種類の問題セット全体を通して，以下の学修項目に関連する設問が頻繁に見られた。

#### ・データの視覚的読解（項目2-2-1）

散布図，箱ひげ図，ヒストグラム，折れ線グラフ，棒グラフ，帯グラフ，さらにはこれらを組み合わせた散布図行列（項目2-1-9）など，多様なグラフを正確に読み取り，データの特徴や傾向を把握する能力が繰り返し問われている。

#### ・統計的指標の理解と活用（項目2-1-2，2-1-4，2-1-5）

代表値（中央値，平均値），ばらつき（四分位数，標準偏差），相関係数といった基本的な統計量の意味を理解し，データの分布比較や変数間関係の解釈に適用する能力が求められている。

#### ・相関と回帰の基礎（項目2-1-5，4-8-4）

変数間の相関関係を散布図や相関係数から読み取り，その意味を解釈する問題が多く見られる。また，単回帰直線を用いた予測値や残差の計算・解釈に関する設問も複数のセットで出題されている。

#### ・層別，比較思考（項目2-1-7，2-2-2）

特定の条件に基づいてデータをグループ分けし，グループ間の特徴や傾向の違いをグラフや統計量から比較・考察する能力を問う問題が出題されている。

#### ・時系列データの基礎（項目4-4-1，4-4-2）

参考問題においては、時系列データの基本的な変動要素（周期性）や移動平均といった概念が問われた。

これまでの共通テスト「情報Ⅰ」では、データリテラシーの基礎として、特にデータやグラフを正確に読み解く能力、および統計的な思考力を問う問題が主に扱われていると言える。

## 5.2 共通テストでの出題が限定的な項目

一方で、モデルカリキュラムに含まれるものの、今回の分析対象では出題が見られないか、限定的であった項目も存在する。

### ・データハンドリングスキル（項目 2-2-3, 4-7 等）

データの取得方法（項目 2-3）、データハンドリングやプログラミング言語を用いたデータ処理（項目 4-7）など、実践的なデータ操作スキルを直接問う設問はほとんど見られなかった。スプレッドシート等の利用は背景として想定されているが、その操作自体が評価対象ではない。

### ・因果関係の深い理解（項目 2-1-5）

相関関係、相関と因果の区別は選択肢で問われているものの、より深い因果推論に関する考察を求める設問は限定的であった。

### ・高度な分析手法（オプショナル項目）

重回帰分析、ロジスティック回帰分析（項目 4-8-4）、クラスタリング（項目 4-9）

重回帰分析やロジスティック回帰分析、クラスタリングのようなより発展的な分析手法に関する出題は見られなかった。

### ・情報伝達スキル、プレゼンテーションスキル（項目 2-5-5）

情報伝達スキルやプレゼンテーションスキルなどは、ペーパーテストという形式上の制約もあり、出題が難しいと考えられる。

## 5.3 高大接続への示唆

「情報Ⅰ」の「データの活用」分野で育成される能力は、大学における DS・AI 教育（リテラシーレベル）の重要な基盤となることは明らかである。多様なグラフや統計量に触れ、データから情報を読み取る経験は、より高度なデータサイエンスを学ぶ上で不可欠な素養である。しかし、大学でより本格的なデータ分析に取り組むためには、高校段階での学習だけでは不十分な側面もある。特に、実践的なデータハンドリングスキル、統計的推論を含むより深い分析手法、因果関係への理解、データ倫理（ELSI）への意識などは、大学教育において体系的に、かつ発展的に育成していく必要がある。共通テスト「情報Ⅰ」は、これらの大学での学びへの橋渡しとして、基礎的なリテラシーの到達度を確認する役割を担う一方で、大学側は「情報Ⅰ」履修者が持つ知識・スキルを前提としつつ、さらなる深化・拡充を図るカリキュラム設計が求められる。

## 6. おわりに

本研究は、高等学校「情報Ⅰ」における「データの活用」分野に関する学習内容と、大学教育における DS・AI 教育（リテラシーレベル）の標準的な枠組みであるモデルカリキュラムとの接続性を明らかにすることを目的とした。そのために、まず学習指導要領解説とモデルカリキュラムの

内容比較を行い、次いで共通テスト「情報Ⅰ」の関連公開問題 6 種類を対象に、モデルカリキュラムの学修項目との対応関係を分析した。

分析の結果、高等学校「情報Ⅰ」の「データの活用」分野で育成される能力、特にデータの種類や基本的な統計量の理解、多様なグラフ（ヒストグラム、散布図、箱ひげ図等）を読み解く力、関連の概念把握などは、モデルカリキュラムのコア項目である「データを読む」「データを説明する」能力と強い関連性を持つことが確認された。さらに、共通テスト「情報Ⅰ」の「データの活用」分野では、これらの基礎的なデータリテラシー（視覚的読解、統計量解釈、相関理解、層別比較など）が問われていることが明らかになった。これらの結果は、高等学校「情報Ⅰ」が大学での DS・AI 教育の導入・基礎として有効に機能しうるポテンシャルを持つことを示唆している。

一方で、両者の間には接続における課題、すなわちギャップも存在する。モデルカリキュラムで重視される因果推論の深い理解、データベース操作やプログラミングを含む実践的なデータハンドリングスキル、重回帰分析等のより高度な分析手法、そしてデータ利活用に伴う倫理的・法的・社会的課題（ELSI）への深い意識などは、現在公開されている共通テストでの出題では限定的であり、高等学校「情報Ⅰ」の標準的な範囲内としては捉えづらいと考えられる。これらの領域については、大学教育段階において、高等学校での学習内容を前提としつつ、接続的かつ重点的に、あるいは新たに指導する必要がある。

本研究の知見は、高等学校と大学が DS 教育におけるそれぞれの役割と接続性を明確に認識し、連携を強化していく上で基礎資料となるものである。両者が体系的な教育プログラムを構築・改善していくことで、社会で必要とされる DS 素養を持つ人材をより効果的に育成することが期待される。

なお、本研究は公開されている問題セットに基づく分析であり、今後の共通テストの出題傾向の変化や、分析者による解釈の限界も存在する。また、本研究は共通必修科目「情報Ⅰ」との対応であるが、選択科目「情報Ⅱ」との対応関係は示していない。

今後の課題としては、実際の共通テスト問題の継続的な分析、および高等学校「情報Ⅰ」の教科書との関連性や、授業実践と生徒の学習成果に関する実証的な調査を通じて、データサイエンス教育における高大接続のあり方をさらに深く検討していくことが挙げられる。

### 謝辞

本研究の一部は JSPS 科研費 JP23K02665 の助成を受けている。

### 参考文献

- [1] 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～”, 入手先〈[http://www.mi-u-tokyo.ac.jp/consortium/model\\_literacy.html](http://www.mi-u-tokyo.ac.jp/consortium/model_literacy.html)〉（参照 2025-05-05）.
- [2] 文部科学省, “高等学校学習指導要領（平成 30 年告示）解説情報編”（2018）.
- [3] 文部科学省, “高大接続システム改革会議最終報告”（2022）.
- [4] 文部科学省, “新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について（答申）”, 中央教育審議会（2014）.

- [5] 村松浩幸, “縦横連携で小中高大一貫の情報教育実現を”, 情報処理, Vol.62, No.7, p.325(2021).
- [6] 小野永貴, “大学の一般情報教育と接続・統合した情報リテラシー教育～小中高で一貫した情報リテラシー教育への課題～, 情報の科学と技術, Vol.67, No.10, pp.539-545(2017).
- [7] 文部科学省, “【資料 2-1】大学入学共通テストへの『情報 I』の導入について”, 入手先 [https://www.mext.go.jp/content/20211021-mxt\\_daigakuc02-000018569\\_3.pdf](https://www.mext.go.jp/content/20211021-mxt_daigakuc02-000018569_3.pdf) (参照 2025-05-05).
- [8] 吉田翔太郎, “大学入学共通試験への教科「情報」追加の経緯に関する考察—2010年代初頭以降の動向に着目して—”, 大学入試研究ジャーナル, Vol.34, pp.52-59(2024).
- [9] 清水克彦, “普通教科「情報」とデータサイエンス”, Science Forum, Vol.42, pp.58-59(2025).
- [10] 大橋 真也, “高等学校共通教科「情報」におけるデータサイエンス”, コンピュータ & エデュケーション, Vol.52, pp.18-23(2022).
- [11] 市川隆司, “高等学校情報科「情報 I」教科書におけるデータサイエンスの要素に関する考察”, 大阪信愛学院短期大学紀要, Vol.57:B1(2022).
- [12] 井手広康, “情報 I の教科書におけるプログラミング分野の比較と考察”, 情報処理学会論文誌教育とコンピュータ, Vol.8, No.3, pp.8-18(2022).
- [13] 大学入試センター, “平成 30 年公示高等学校学習指導要領に対応した大学入学共通テストの『情報』の試作問題（検討用イメージ）” 入手先 <https://www.ipsj.or.jp/education/edu202012.html> (参照 2025-05-05).
- [14] 大学入試センター, “サンプル問題『情報』問題”, 入手先 [https://www.dnc.ac.jp/kyotsu/shiken\\_jouhou/r7/r7\\_kentoujoukyou/](https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7/r7_kentoujoukyou/) (参照 2025-05-05).
- [15] 大学入試センター, “試作問題『情報 I』”, 入手先 [https://www.dnc.ac.jp/kyotsu/shiken\\_jouhou/r7/r7\\_kentoujoukyou/r7mondai.html](https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7/r7_kentoujoukyou/r7mondai.html) (参照 2025-05-05).
- [16] 大学入試センター, “試作問題『情報 I』（参考問題）”, 入手先 [https://www.dnc.ac.jp/kyotsu/shiken\\_jouhou/r7/r7\\_kentoujoukyou/r7mondai.html](https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7/r7_kentoujoukyou/r7mondai.html) (参照 2025-05-05).
- [17] 朝日新聞, “2025 年度大学入学共通テスト 情報 I 問題全文”, 入手先 [https://www.asahi.com/edu/kyotsu-exam/shiken2025/mondai\\_day2\\_5tuepw7sd/joho1.html](https://www.asahi.com/edu/kyotsu-exam/shiken2025/mondai_day2_5tuepw7sd/joho1.html) (参照 2025-05-05).
- [18] 福島民友みんゆう Net, “2025 年度大学入学共通テスト追・再試験問題”, 入手先 <https://www.minyu-net.com/common/files/2025joho1.pdf> (参照 2025-05-05).
- [19] 大学入試センター, “サンプル問題『情報』ねらい”, 入手先 [https://www.dnc.ac.jp/kyotsu/shiken\\_jouhou/r7/r7\\_kentoujoukyou/#anchor02](https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7/r7_kentoujoukyou/#anchor02) (参照 2025-05-05).
- [20] 大学入試センター, “試作問題及び試作問題の概要等概要「情報」”, 入手先 [https://www.dnc.ac.jp/kyotsu/shiken\\_jouhou/r7/r7\\_kentoujoukyou/r7mondai.html](https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7/r7_kentoujoukyou/r7mondai.html) (参照 2025-05-05).

付録 A モデルカリキュラム（データリテラシー）におけるデータ分析関連キーワード（データリテラシーの抜粋）

| 項目            | キーワード（知識・スキル）                                                                                                                                                                                                                                                                                                                                                                            |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2-1. データを読む   | 2-1-1. データの種類（量的変数, 質的変数）<br>2-1-2. データの分布（ヒストグラム）と代表値（平均値, 中央値, 最頻値）<br>2-1-3. 代表値の性質の違い（実社会では平均値＝最頻値でないことが多い）<br>2-1-4. データのばらつき（分散, 標準偏差, 偏差値）, 外れ値<br>2-1-5. 相関と因果（相関係数, 擬似相関, 交絡）<br>2-1-6. 観測データに含まれる誤差の扱い<br>2-1-7. 打ち切りや欠測を含むデータ, 層別の必要なデータ<br>2-1-8. 母集団と標本抽出（国勢調査, アンケート調査, 全数調査, 単純無作為抽出, 層別抽出, 多段抽出）<br>2-1-9. クロス集計表, 分割表, 相関係数行列, 散布図行列<br>2-1-10. 統計情報の正しい理解（誇張表現に惑わされない） |
| 2-2. データを説明する | 2-2-1. データ表現（棒グラフ, 折線グラフ, 散布図, ヒートマップ, 箱ひげ図）<br>2-2-2. データの比較（条件をそろえた比較, 処理の前後での比較, A/B テスト）<br>2-2-3. 不適切なグラフ表現（チャートジャンク, 不必要な視覚的要素）<br>2-2-4. 優れた可視化事例の紹介（可視化することによって新たな気づきがあった事例など）<br>2-2-5. 相手に的確かつ正確に情報を伝える技術や考え方（スライド作成, プレゼンテーションなど）                                                                                                                                             |
| 2-3. データを扱う   | ・データの取得（機械判読可能なデータの作成・表記方法）<br>・データの集計（和, 平均）<br>・データの並び替え, ランキング<br>・データ解析ツール（スプレッドシート, BI ツール）<br>・表形式のデータ（csv）                                                                                                                                                                                                                                                                        |

付録 B モデルカリキュラム（データリテラシー）におけるデータ分析関連キーワード（オプションの抜粋）

| オプション                | キーワード（知識・スキル）                                                                                                                                                                                                                  |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4-4. 時系列データ解析        | 4-4-1. 時系列データ（トレンド, 周期, ノイズ）<br>4-4-2. 季節調整, 移動平均                                                                                                                                                                              |
| 4-8. データ活用実践（教師あり学習） | 4-8-1. 教師あり学習による予測（例）売上予測, 罹患予測, 成約予測, 離反予測など<br>4-8-2. データの収集（分析に必要なデータの確認, 対象となるデータの収集）<br>4-8-3. データの加工（データクレンジング, サンプリング, 簡単な説明変数の作成）<br>4-8-4. データの分析（単回帰分析, 重回帰分析, ロジスティック回帰分析, モデルの評価）<br>4-8-5. データ分析結果の共有, 課題解決に向けた提案 |