

# 大学生を対象としたデータサイエンスに関する 数学的基礎能力の測定とニーズ及び影響要因の調査

辻 靖彦<sup>1)</sup>、秋葉広人<sup>2)</sup>、芝崎順司<sup>3)</sup>

## Survey on Mathematical Basic Ability by Testing Related to Data Science, Its factors and Needs Targeting at Universities' Students

Yasuhiko TSUJI, Hiroto AKIBA, Junji SHIBASAKI

### 要 旨

放送大学で配信中のデータサイエンス (DS) のモデルカリキュラム準拠のリテラシー教材を利用した学習が十分に効果を上げるためには、その学習内容を理解するために必要な高等学校卒業レベルの「数学」や「情報」科目で扱う学習内容を修得していることが前提となる。そこで本論文では、データサイエンス教育における基礎学力の把握及び影響要因とニーズ調査のために、データサイエンス分野のリメディアルチェックテストとアンケートを開発し、2つの大学の1年生計140名を対象に実施した。その結果、DSに関連する高校数学のリメディアル分野において学生がつまづきやすい事項が示唆されると共に、DSに関連する高校数学の科目や単元の受講が一部に留まっている現状、そしてDSスキルへのニーズとDS関連用語の理解度の実情が明らかになった。さらに、テストの正答率と調査項目の関連性を調べた結果、数学Ⅱと数学B、そしてDSに関連する数学Ⅰの「データの分析」、数学Aの「場合の数と確率」、数学Bの「確率分布と統計的な推測」の単元を履修した学生、そしてプレゼンテーションソフトと表計算ソフトを学校で習った学生の正答率が有意に高く、これらの事項がDSに関する数学的基礎能力の向上と関連のある可能性が示唆された。

**キーワード**：データサイエンス、リメディアル教育、数学教育、情報教育

### ABSTRACT

In order to learn effectively using literacy teaching materials based on the data science model curriculum being distributed at the Open University of Japan, high school graduation-level “math” and “information” are required to understand the learning content. It is assumed that you have mastered the learning content to be dealt with in the subject. Therefore, in this paper, we developed a remedial check test and a questionnaire in the field of data science in order to understand basic academic ability in data science education and to investigate influential factors and needs, and conducted it for a total of 140 first-year students from two universities. As a result, it is suggested that students are likely to trip over in the field of high school mathematics remedial related to DS, and the current situation that only some high school mathematics subjects and units related to DS are taken, and DS skills. The actual situation of needs and understanding of DS-related terms was clarified. Furthermore, as a result of investigating the relationship between the correct answer rate of the test and the survey items, “Analysis of data” of Mathematics II and Mathematics B, and DS-related Mathematics I, “Number and probability of cases” of Mathematics A, and Mathematics B. The percentage of correct answers for students who took the unit of “Probability Distribution and Statistical Guess” and those who learned presentation software and table calculation software at school was significantly high, and these items improved the mathematical basic ability regarding DS.

**Key words** : data science, remedial education, mathematics education, information education

<sup>1)</sup> 放送大学准教授 (「情報」コース)

<sup>3)</sup> 放送大学教授 (「情報」コース)

<sup>2)</sup> 北里大学保健衛生専門学院

## 1. 背景と目的

我が国ではAI戦略2019において、2025年までに具体的目標として「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」することが定められている。それに呼応する形で現在、放送大学ではBS231ch.とインターネット配信公開講座による、データサイエンスのモデルカリキュラムに準拠したリテラシー教材（以下、DSリテラシー教材）を制作し、配信を行っている（放送大学 2021）。

このようなDSリテラシー教材を用いた学習が十分な効果を上げるためには、その学習内容を理解するために必要な高校卒業レベルの「数学」や「情報」科目で扱う学習内容について修得していることが前提となる。しかし、大学生を対象とした調査によると数学の基礎事項を理解出来ていない学生が一定数存在することが指摘されており（日本数学会 2012、芳沢 2019など）、特にDSリテラシー教材の基礎に当たる「データを読む、説明する、扱う」に関する基礎学力が不十分な学生が多いという結果が出ている。また、学生によってはDSリテラシー教材を利用した授業で扱う内容について、高校までの授業で基礎的内容を十分に学習していない範囲が含まれる。

そこで本研究では、DSリテラシー教材を利用した学習を効果的にすることを最終目的とし、DSリテラシー教材の内容に関連する高等学校以前の学習内容の習得の実態を把握することを研究目的としたリメディアルチェックテストを開発した。そして複数の大学に所属する大学1年生を対象にテストを実施すると共に、DS能力との関連性を調べるために高校「数学」及び教科「情報」の履修状況やDSスキルに対するニーズや習得状況、そしてDS関連の用語理解度等に関するアンケート調査を行った。

## 2. 研究方法

### 2.1 データサイエンスリメディアルチェックテストの開発

テストの開発のために、筆者らは始めに小・中・高等学校の現行の学習指導要領算数編・数学編を基に、DSに関連する用語のリストアップを行い、初等中等教育におけるDSに関連する内容の確認を行った。その上で問題の学年レベルの難易度を設問単位で把握できるようにするために、実用数学技能検定（日本数学検定協会 2021）及び統計検定（統計質保証推進協会 2021）に用いられているDS関連の問題を基に、小学2年～高校2年レベルの全20問を3セット作成し、WebClassのLMS上に実装した。テスト時間は60分間と設定し、問題の提示順はランダムとした。3セットの中の2セットを事前と事後のリメディアルチェックテスト、残りの1セットを演習用の練習問題とした。

### 2.2 アンケート項目の設計

本アンケートはデータサイエンスに関する数学的基礎能力、つまりチェックテストの正答率との関連性やDSスキルへのニーズを調べることを目的としている。調査項目は「高校数学の履修状況」、「小学・中学・高校における情報教育」、「現在の状況」、「用語理解度チェック」の4つのカテゴリに分かれている（表1）。いずれもDSの基礎学力に与える影響として考えられる項目を、先行研究を基に取り上げた。

具体的にどのように設計したのかについて述べる。「高校数学の履修状況」では高校数学における科目や単元における履修状況と、高校の数学の授業に満足していたか、好きだったか、そして得意だったかどうかを聞いている。高校数学における履修状況は古川（2017）を基に、また、高校数学の満足度は河村（2019）が教科「情報」への満足度を調査しており、それを高校数学へ転用する形で設計した。

「小学・中学・高校における情報教育」では高校における教科「情報」の履修状況とこれまで受けた情報教育の内容、そして数学と同様に教科「情報」の授業の満足度、好きだったかどうか、得意だったかどうかを調査する。これらはいずれも河村（2019）を基に設計した。

「現在の状況」では現在の調査対象者のICTスキルについて聞いている。具体的には、コンピュータに対する印象（河村 2019を基に修正）およびデータサイエンススキルの習得状況及びニーズについて竹内・末永（2018）のデータサイエンススキル・能力を引用して設計した。

「用語理解度チェック」では筆者らが小学校・中学校・高等学校の学習指導要領算数編と数学編をもとに、データサイエンスに関連すると考えられる用語のリストアップを行い、42語を抽出した。そして木村・伊澤（2016）の調査方法を参考に「理解している」

表1 調査項目

カテゴリ	調査項目
高校数学の履修状況 <sup>*1</sup>	履修した高校数学の授業科目、受講した単元、授業の満足度、授業の印象（好き／嫌い）、授業への意識（得意／苦手）
小学・中学・高校における情報教育 <sup>*2</sup>	教科「情報」の履修状況、コンピュータを使い始めた時期、情報教育の学習内容、授業の満足度、授業の印象（好き／嫌い）、授業への印象（得意／苦手）
現在の状況	コンピュータへの印象 <sup>*2</sup> 、DSスキルをどの程度身に付けているか／在学中に身に付けたいと思うか <sup>*3</sup>
用語理解度チェック	初等中等教育の数学で学習するDS関連用語をどの位、理解しているか <sup>*4</sup>

<sup>\*1</sup> 古川（2017）を参考に作成 <sup>\*2</sup> 河村（2019）を基に作成

<sup>\*3</sup> 竹内・末永（2018）のDSスキル・能力を引用 <sup>\*4</sup> 木村・伊澤（2016）を基に作成

「ある程度理解している」「何も分からない」の3件法で各用語に対する理解度を尋ねた。

なお、本アンケートもリメディアルチェックテストと同様にWebClassのLMS上に実装した。また、リストアップした42の用語においてはその説明教材を合わせて作成した。これは事前テストと事後テストの間に学習を挟むことでDSリメディアルの学習を促す目的で制作した。

### 3. 大学生を対象としたDSリメディアルチェックテスト及びアンケート調査の実施

#### 3.1 概要

私立のA大学の文系学部に所属する大学1年生102名および、私立のB大学の医療系学部に所属する大学1年生79名を対象に、開発したリメディアルチェックテスト及びアンケート調査を実施した。調査期間はA大学が2021年4月16日～6月22日、B大学が2021年5月12日～5月21日である。実施方法について述べる。A大学とB大学のいずれも、始めに一度対面環境にてテストの受験方法とアンケートの回答方法について説明し、その後、学習者は期間中に自宅等の任意の場所から任意の時間にインターネット経由でLMSにアクセスして受験や回答を行った。

#### 3.2 結果

本論文では、リメディアルチェックテストおよびアンケートの基礎集計、そしてチェックテストの正答率と調査項目の関連性について分析した。表2に各テスト、アンケート、そして教材の利用者数を示す。これより、DSリメディアルチェックテストの利用人数は137名、アンケートの回答数は135～137名であった。なお、練習問題、用語解説、リメディアルチェックテスト（事後）はA大学とB大学のいずれにおいても学習を必須としなかったため利用者数が少なかった結果となった。

##### 3.2.1 テストの正答率

DSリメディアルチェックテストの平均正答率は0.528、標準偏差は0.210であった。正答率のヒストグラムを図1に示す。さらに、各問題の難易度と正答率

を図2に示す。図2では各問題を難易度の易しい順に並べており、右側に行くほど難易度が上がる。これより各問題の正答率を比較したところ、

- 日付のカウント（問題番号4）
- 四捨五入と概数（問題番号5）
- クロス集計表（問題番号6）
- 帯グラフから割合を読み取る問題（問題番号7）
- 長文で思考力を問う問題（問題番号13、18）
- 複数のサイコロを扱った確率計算（問題番号14、17）
- 度数分布表と相対度数（問題番号15）
- 累積相対度数（問題番号19）
- 度数分布表に関する問題（問題番号20）

において正答率が4.4～38.5%と低い傾向がみられた。難易度と正答率を比較すると、日付のカウント、四捨五入と概数、クロス集計表、帯グラフから割合を読み取る問題はいずれも小学3年～5年に習得する内容であるにも係わらず正答率は低い傾向がみられた。また、高校1年レベルの度数分布表と相対度数（問題番号15）の問題においてはおおむね計算は合っているものの「小数第3位を四捨五入して小数第2位まで表す」や「小数第2位を四捨五入して小数第1位まで求める」といった四捨五入のケアレスミスが多くみられた。この傾向は四捨五入と概数の問題（問題番号5）においても同様にみられた。

##### 3.2.2 調査結果①—高校数学の履修状況

高校数学の履修状況の回答結果を図3に示す。これより、必修である数学Ⅰに加えて、8割以上の学生が高校時代に数学Aと数学Ⅱを履修したことが分かった。その一方で数学活用を履修した学生はおらず、数学Bは40.9%、数学Ⅲは14.6%の学生のみの履修に留まっていた。各科目を履修した学生に対して単元別に受講の有無を聞いた結果を表3に示す。これより、数学Aの「場合の数と確率」は82.8%と多くの学生が受講したと回答したものの、数学Ⅰの「データの分析」は53.7%、数学Bの「確率分布と統計的な推測」は62.5%と、DSに関連すると考えられる単元において一部の生徒の受講に留まっていることが分かった。特

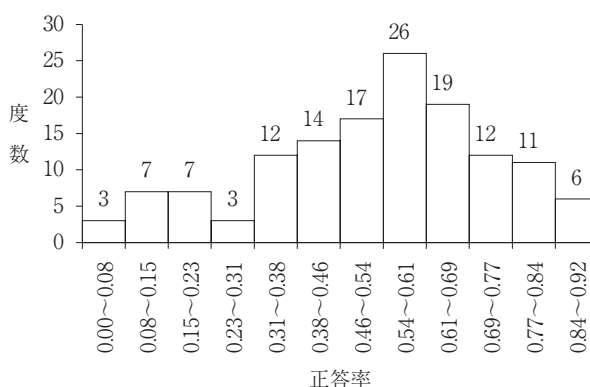


図1 DSリメディアルチェックテストの正答率のヒストグラム ( $n=137$ )

表2 各教材の利用者数

教材名	利用者数
リメディアルチェックテスト	137
アンケート①—高校数学の履修状況	137
アンケート②—小学・中学・高校における情報教育	136
アンケート③—現在の状況	135
アンケート④—用語理解度チェック	136
練習問題	63
用語解説	59
リメディアルチェックテスト（事後）	57
合計	140

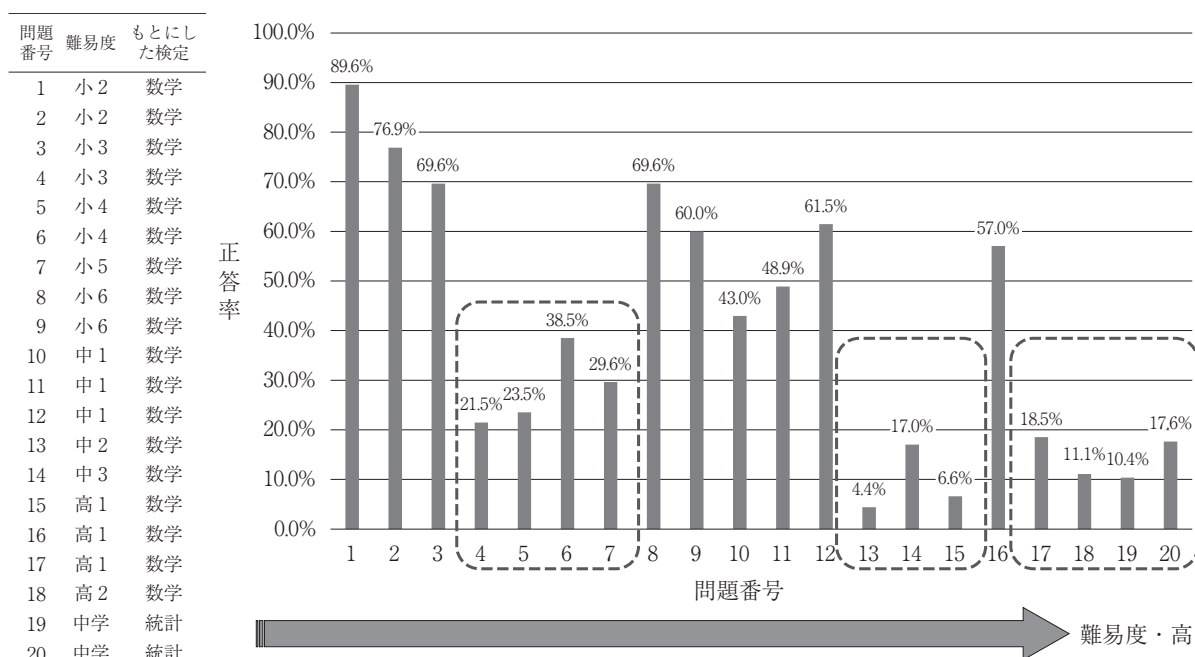


図2 難易度別の各問題の正答率

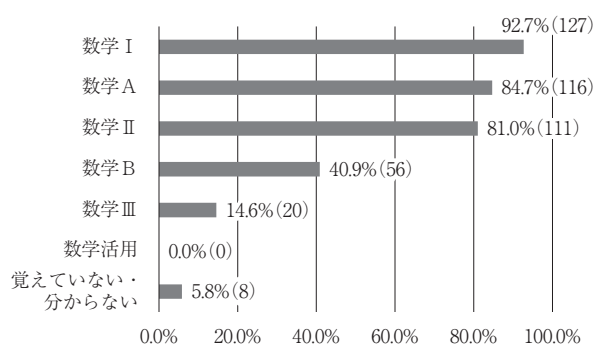


図3 高校数学の履修状況 (n=137)

に数学Bに関しては科目自体の履修率が約4割と低いことから、「確率分布と統計的な推測」の単元の履修率は母数を全体と考えると約25%と一部に留まっていることが分かる。

また、高校の数学の授業の満足度、好きか嫌い、そして得意だったかどうかの回答結果を図4に示す。これより、約40%の生徒が数学の授業に（どちらかと言えばを含め）満足しており、約35%の生徒が数学の授業が（どちらかと言えばを含め）好き、もしくは得意だったことが伺える。しかしその一方で、約3割の生徒が「数学が嫌い／苦手」と回答していた。これらはいずれも大学生の実情を表していると考えられる。

### 3.2.3 調査結果②—小学・中学・高校における情報教育

本節では、小学・中学・高校における情報教育として、はじめに高校における教科「情報」の履修状況、情報の授業の満足度・好き嫌い・得意不得意、そして小学・中学・高等学校で学習した情報教育の内容と基

表3 数学の各単元の受講状況

科目	単元名	受講率(度数)
数学I (n=137)	1. 数と式	81.6% (111)
	2. 図形と計量	72.1% (98)
	3. 二次関数	83.8% (114)
	4. データの分析	53.7% (73)
	5. どの単元を受講したか覚えていない	18.4% (25)
数学A (n=136)	1. 場合の数と確率	82.8% (96)
	2. 整数の性質	69.8% (81)
	3. 図形の性質	67.2% (78)
	4. どの単元を受講したか覚えていない	25.9% (30)
数学II (n=111)	1. いろいろな式	76.6% (85)
	2. 図形と方程式	75.7% (84)
	3. 指数関数・対数関数	78.4% (87)
	4. 三角関数	77.5% (86)
	5. 微分・積分の考え	81.1% (90)
	6. どの単元を受講したか覚えていない	19.8% (22)
数学B (n=56)	1. 確率分布と統計的な推測	62.5% (35)
	2. 数列	92.9% (52)
	3. ベクトル	89.3% (50)
	4. どの単元を受講したか覚えていない	19.6% (11)
数学III (n=20)	1. 平面上の曲線と複素数平面	55.0% (11)
	2. 極限	70.0% (14)
	3. 微分法	95.0% (19)
	4. 積分法	100.0% (20)
	5. どの単元を受講したか覚えていない	25.0% (5)

礎的スキルについて述べる。続いて調査対象者のICTスキルを調べるために、コンピュータを使い始めた時期に関する回答結果を述べる。

教科「情報」の履修状況を表4に示す。これより、7割を超える学生が「社会と情報」と「情報の科学」のいずれか、もしくは2科目両方の科目を履修したことが分かる。また、情報の授業の満足度、好きか嫌い



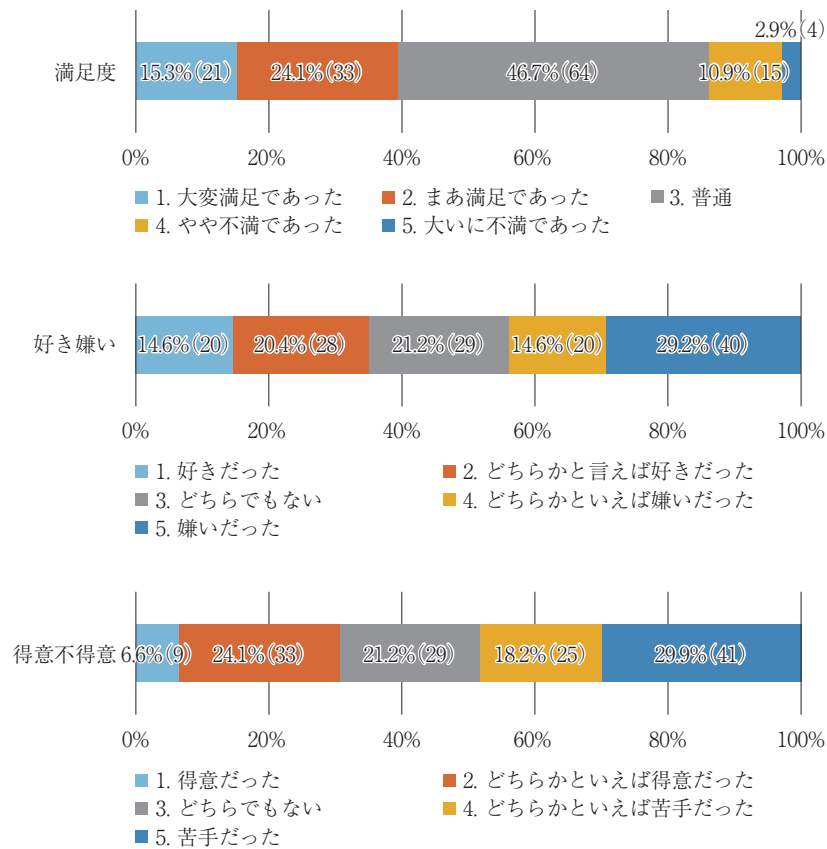


図4 数学の授業の満足度（上）、好き嫌い（中）、得意不得意（下）（ $n = 137$ ）

表4 教科「情報」の履修状況

選択肢	割合(度数)
1.「社会と情報」を履修した	41.2% (56)
2.「情報の科学」を履修した	16.9% (23)
3. 両方履修した	5.1% (7)
4. 覚えていないがどちらか1科目を履修した	11.0% (15)
5. どちらも履修していない	8.1% (11)
6. 覚えていない・分からない	17.6% (24)
合計	100.0% (136)

か、そして得意だったかどうかの回答結果を図5に示す。数学の結果と比べると情報の授業は満足度が若干低い傾向があるものの、好きもしくはどちらかと言えば好きと回答した生徒は数学よりも少し多く、4割を超えていることが分かった。また、嫌いもしくは苦手と回答した生徒も数学と比べるととても少ないことが分かった。高校だけに限定していないが、小学・中学・高校の授業の中で学習した情報教育の内容の回答結果を表5に示す。これを見ると全体として割合が2割以下と低い傾向がみられたものの、最も回答の割合が高い項目は「1. コンピュータを使って文書を作成したりデータを処理する」と「2. インターネットを使って調べる」であった（いずれも16.9%）。この2つの項目は所謂、調べ学習の中で学んだ事項と考えられる。次に高かった項目は「3. コンピュータやイン

表5 小・中・高校の授業で学んだ情報教育の内容（複数回答可）

選択肢	割合(度数)
1. コンピュータを使って文書を作成したりデータを処理する	16.9% (23)
2. インターネットを使って調べる	16.9% (23)
3. コンピュータやインターネットのしくみ	11.0% (15)
4. 身近なデータベースシステムの活用事例	4.4% (6)
5. コンピュータを用いたモデル化やシミュレーション	3.7% (5)
6. 情報セキュリティ	5.9% (8)
7. ICTを使って起きている社会的問題と法やモラル	0.7% (1)
8. 覚えていない	1.5% (2)
9. 履修していない	0.0% (0)
10. その他	3.7% (5)
合計	100.0% (136)

ターネットのしくみ」(11.0%)であった。また、授業の中で習った基礎的スキルの回答を表6に示す。これより、「1. キー入力（タイピング）」が最も回答の割合が84.6%と高く、続いて「2. ワードプロ（Word等）」(69.1%)、「3. 表計算（Excel等）」(60.3%)、「4. プレゼンテーション（PowerPoint等）」(60.3%)となっていた。これより、基礎的なICTスキルは授業の中で学習したことが窺えた。その一方でWeb作成

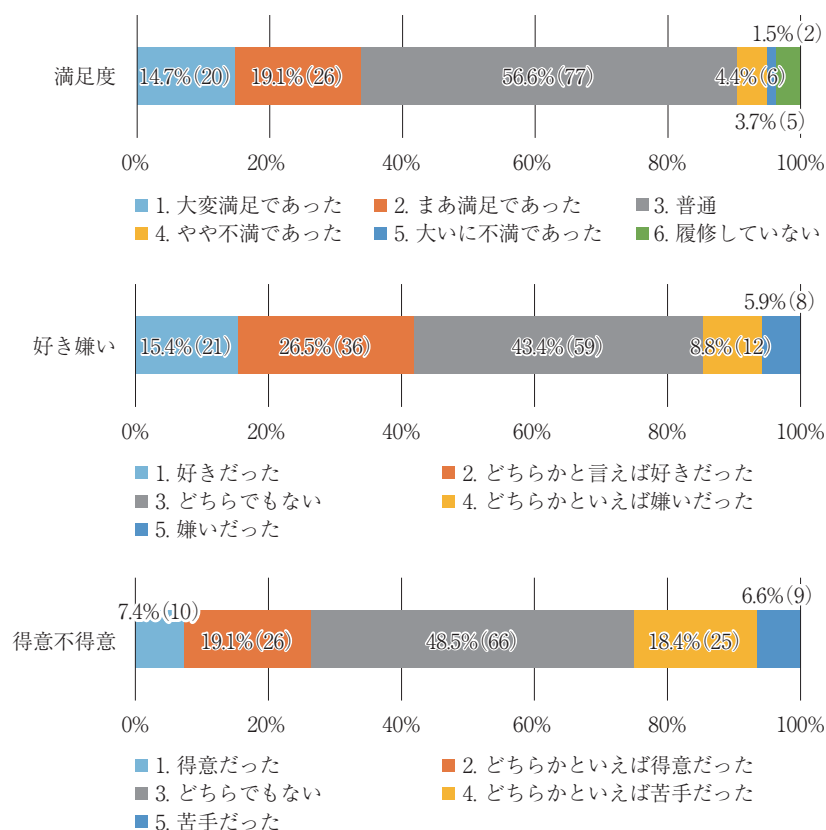
図5 情報の授業の満足度（上）、好き嫌い（中）、得意不得意（下）（ $n=136$ ）

表6 小学・中学・高校の授業で習った基礎的スキル（複数回答可）

選択肢	割合(度数)
1. キー入力（タイピング）	84.6% (115)
2. ワードプロ（Word等）	69.1% (94)
3. 表計算（Excel等）	60.3% (82)
4. プレゼンテーション（PowerPoint等）	60.3% (82)
5. Web作成	12.5% (17)
6. プログラミング	14.7% (20)
7. 該当なし	5.9% (8)
合計	100.0% (136)

やプログラミングといったやや高度なスキルを習った学生はいずれも15%未満と少なかった。

続いて、学生のICTスキルに関する項目の回答結果について述べる。コンピュータを使い始めた時期は、学校以外で使い始めたと回答した学生は高校入学以前・以後合わせても約25%と少なく、約7割を超える学生が学校でコンピュータを使い始めたことが分かった（表7）。

### 3.2.4 調査結果③—現在の状況

本節では回答者の現在のDSスキルに関する項目として、コンピュータに対する印象、DSスキルを現在身に付けているか、そしてDSスキルを大学在学中に身に付けていきたいかどうかに関する回答結果を述べる。

表7 コンピュータを使い始めた時期

選択肢	割合(度数)
1. 高校入学時以前に学校で	48.5% (66)
2. 高校入学以前に学校以外で、	21.3% (29)
3. 高校入学以後に学校で、	23.5% (32)
4. 高校入学以後に学校以外で、	2.9% (4)
5. これまで使っていない	3.7% (5)
合計	100.0% (136)

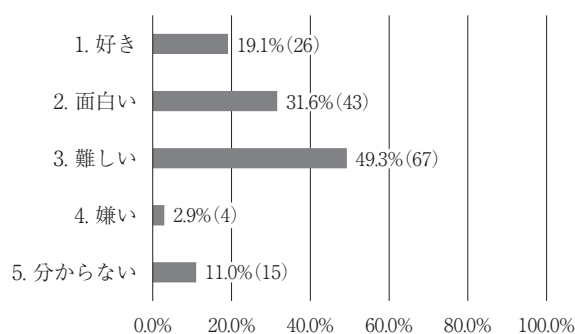
図6 現在のコンピュータに対する印象（ $n=136$ 、複数回答可）

図6に現在のコンピュータに対する印象の回答結果を示す。これより、約2割の学生が「好き」、そして約3割の学生が「面白い」という印象を持っている一方で、約半数の学生が「難しい」という印象を持って

いることが分かった。なお、「嫌い」と回答する学生は2.9%とほとんどいなかった。

DSスキルを現在、どの程度身に付けているか、に関する回答結果を図7に示す。これより、「データ・資料を収集する能力」と「グラフや表の数値を読み取る能力」についてはある程度も含めると6割前後の学生が「身に付けている」と回答した。続いて「分析結果を人に伝える能力（コミュニケーション・プレゼンテーション能力）」が4割弱、「パソコンの表計算ソフト等を使い、簡単なデータ集計や分析をする能力」が3割強の学生が（ある程度も含めて）身に付けている

と回答した。

続いてDSスキル・能力に対するニーズとして、大学在学中にDSスキル・能力を身に付けたいかどうかを聴取した。その回答結果を図8に示す。これより、いずれのスキル・能力に対しても「3. 特に身に付けたくない」と回答した学生はほとんどいなかったことから、DSのスキル・能力に対しての基本的なニーズの高さが窺えた。「1. 専門的なレベルまで身に付けたい」の回答が最も多かったのは「分析結果を人に伝える能力（コミュニケーション・プレゼンテーション能力）」（36.3%）、「データ・資料を収集する能力」（24.4

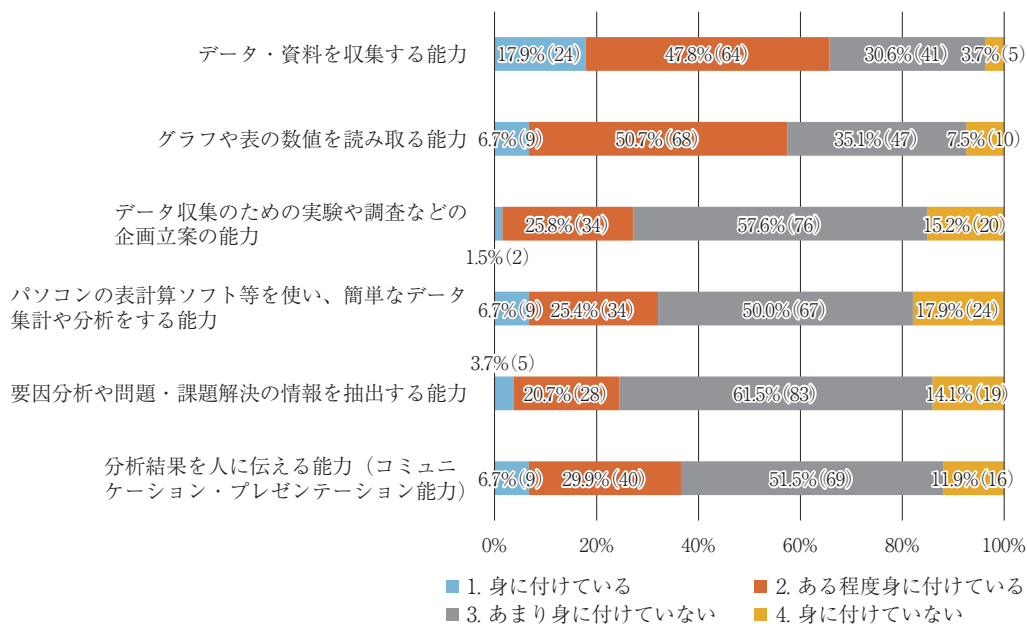


図7 DSスキル・能力をどの程度身に付けているか (n=132~135)

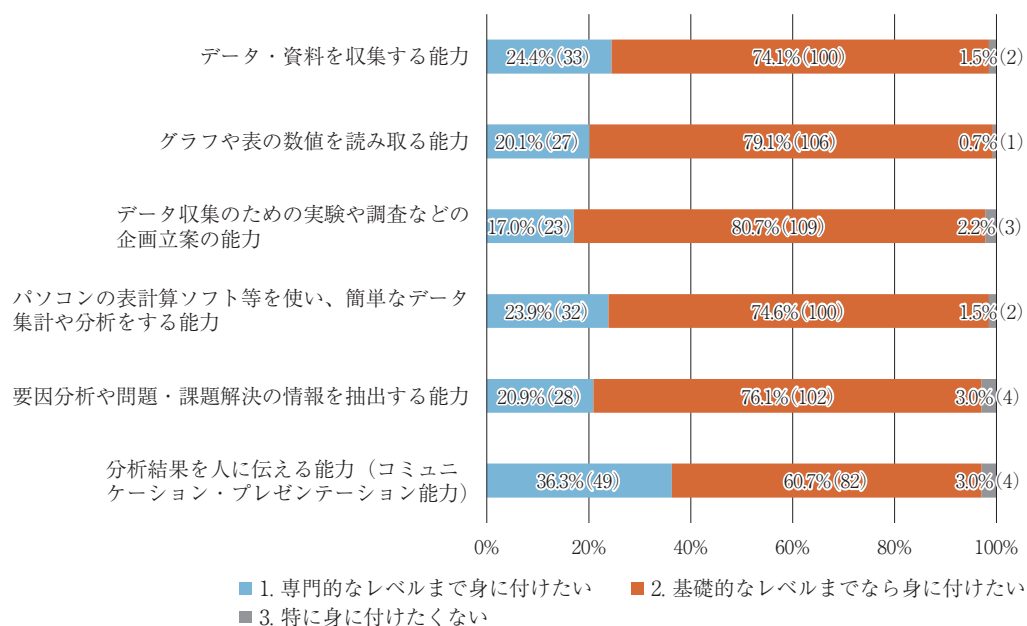


図8 DSスキル・能力を大学在学中に身に付けたいか (n=134~135)

%)、「パソコンの表計算ソフト等を使い、簡単なデータ集計や分析をする能力」(23.9%)の順であった。

### 3.2.5 調査結果④—用語理解度チェック

学習指導要領から抽出したDSに関連する42用語に対する理解度の回答結果を示す(図9)。これより平均値やグラフ、度数などDSの基礎的な事項については多くの割合の学生が理解していることが窺える。しかしその一方で四分位数、四分位範囲、四分位偏差の理解度が低く、また、排反、確率変数、確率分布、二項分布、正規分布など、数学Bの確率統計分野の用語を理解していない学生も多くみられた。

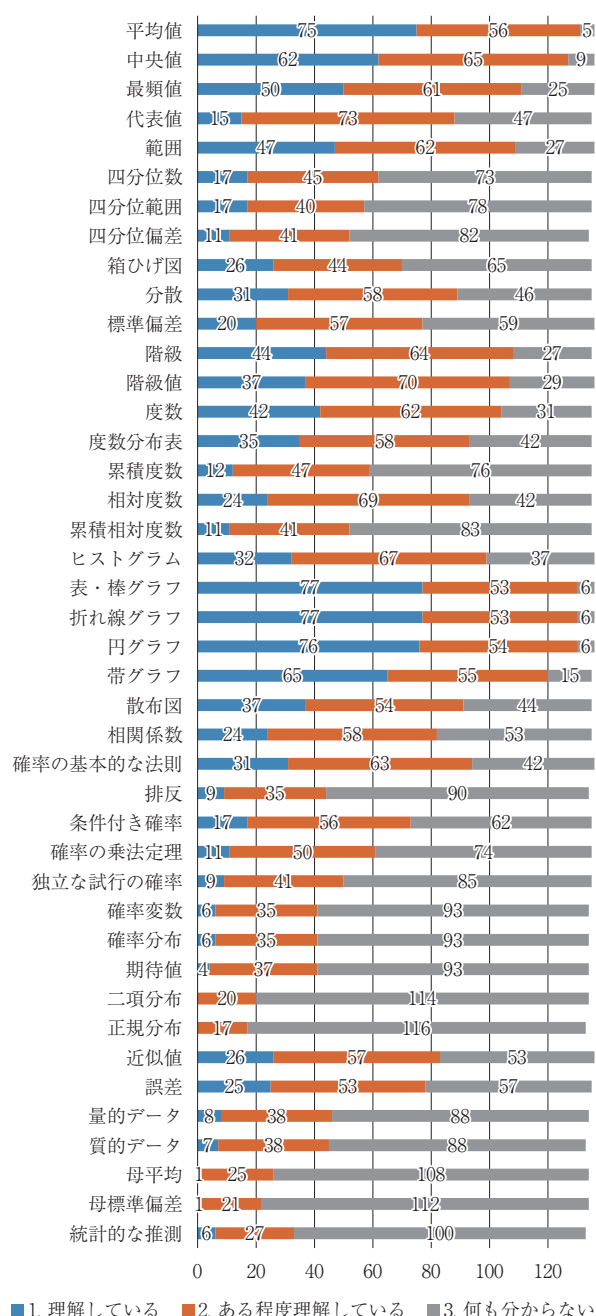


図9 DSの用語理解度 (n=136)

### 3.2.6 テスト正答率と調査項目の関連性

本項では、データサイエンスに関する数学的基礎能力に影響を与える要因を明らかにするために、リメディアルチェックテストの正答率とアンケートの回答を用いて、以下の仮説に対して統計分析を用いて検証した。

仮説(1). 高校数学の各科目及びDSに関連する各単元を履修した学生は、履修していない学生よりもチェックテストの正答率が高い

仮説(2). 教科「情報」を履修した学生は履修していない学生よりもチェックテストの正答率が高い

仮説(3). 高校数学が好きだった／授業に満足していた／得意だった学生はチェックテストの正答率が高い

仮説(4). 教科「情報」が好きだった／授業に満足していた／得意だった学生はチェックテストの正答率が高い

仮説(5). 小学・中学・高校の授業で「表計算 (Excel)」などの基礎スキルを習得した学生はチェックテストの正答率が高い

仮説(6). コンピュータを使い始めた時期が早い学生はチェックテストの正答率が高い

仮説(7). データサイエンスの各スキルを「身に付けている」と認識している学生はチェックテストの正答率が高い

仮説(8). 様々なDS用語を理解している学生はチェックテストの正答率が高い

これらの仮説に対して、1つずつ検証を行った。

仮説(1). 高校数学の各科目及びDSに関連する各単元を履修した学生は、履修していない学生よりもチェックテストの正答率が高い

必修科目に指定されている数学Ⅰを除いた高校数学の各科目(数学A、数学Ⅱ、数学B、数学Ⅲ、数学活用)と、DSに関連する単元として「数学Ⅰ—データの分析」、「数学A—場合の数と確率」、そして「数学B—確率分布と統計的な推測」の3つの各単元に対して履修・受講している学生としていない学生のそれぞれのテストの正答率に対して対応無しのt検定を用いて違いを調べた。その結果を表8に示す。この結果より、科目としては数学Ⅱと数学Bを履修した学生、そして単元ではデータの分析、場合の数と確率、確率分布と統計的な推測をそれぞれ履修した学生は有意に正答率が高かった。これらの科目および単元の受講がDSに関する数学的基礎能力の向上に寄与している可能性が考えられる。

仮説(2). 教科「情報」を履修した学生は履修していない学生よりもチェックテストの正答率が高い

教科「情報」の履修状況(表4)に対して、少なくとも1科目を履修した学生と履修していない・覚えていない学生のそれぞれの正答率を、対応無しのt検定を



表8 数学の履修状況による正答率の違い

科目／単元	履修／受講		未履修／未受講		t 検定結果
	平均正答率 (SD)	n	平均正答率 (SD)	n	
数学A	0.534(0.208)	115	0.497(0.235)	20	$t(133) = -0.73, p = .468, d = -0.175$
数学Ⅱ	0.553(0.197)**	110	0.421(0.241)**	25	$t(133) = -2.88, p = .005, d = -0.635$
数学B	0.576(0.221)*	56	0.494(0.199)*	79	$t(133) = -2.26, p = .025, d = -0.393$
数学Ⅲ	0.512(0.244)	20	0.531(0.206)	115	$t(133) = 0.37, p = .710, d = 0.090$
数学活用	履修者なし				
数学Ⅰ—データの分析	0.587(0.197)**	72	0.467(0.206)**	62	$t(132) = -3.42, p = .001, d = -0.589$
数学A—場合の数と確率	0.549(0.208)*	95	0.467(0.216)*	36	$t(129) = -1.98, p = .050, d = -0.385$
数学B—確率分布と統計的な推測	0.600(0.240)*	35	0.512(0.201)*	73	$t(106) = -1.99, p = .049, d = -0.407$

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

用いて比較した。その結果、有意な違いはみられなかった(表9)。

仮説(3)．高校数学が好きだった／授業に満足していた／得意だった学生はチェックテストの正答率が高い

高校数学の授業の満足度・好き嫌い・得意不得意(図4)の各回答に対して、正答率との間で相関分析を行った。その結果を表10に示す。表10より、数学の授業に対する意識(得意不得意)と正答率の間に有意な弱い正の相関がみられた( $p < .01$ )。これより、数学を得意と認識していた学生は正答率が高い傾向があることが分かった。また、数学の授業の満足度と印象においても有意傾向ではあるがごくわずかな正の相関がみられた( $p < .1$ )

仮説(4)．教科「情報」が好きだった／授業に満足していた／得意だった学生はチェックテストの正答率が高い

高校の教科「情報」の授業の満足度・好き嫌い・得意不得意(図5)の各回答に対して、正答率との間で相関分析を行った。その結果、正答率との間には有意な相関は確認できなかった(表11)。

仮説(5)．小学・中学・高校の授業で「表計算(Excel)」などの基礎スキルを習得した学生はチェックテストの正答率が高い

小学・中学・高校の授業で習った基礎的スキル(表6)に対して、各スキルを習った学生と、習っていない学生における正答率の違いを対応無しt検定を用いて調べた。その結果を表12に示す。これより、「4. プレゼンテーション(PowerPoint等)」と「3. 表計算(Excel等)」を習ったと回答した学生はそうでない学生よりも有意に正答率が高いことが分かった。プレゼンテーションと表計算ソフトの習得意識がDSに関する数学的基礎能力の向上に繋がる可能性が考えられる。また、「6. プログラミング」においても習ったと回答した学生の方が、正答率が高い有意傾向がみられた。

表9 情報の履修状況による正答率の違い

科目／単元	履修／受講		未履修／未受講		t 検定結果
	平均正答率 (SD)	n	平均正答率 (SD)	n	
教科「情報」	0.524(0.217)	99	0.544(0.199)	35	$t(133) = -0.73, p = .468, d = -0.175$

表10 数学の授業の満足度・好き嫌い・得意不得意と正答率の相関

	①	②	③	④
①リメディアルチェックテスト—正答率	1.000			
②高等学校での「数学」の授業の満足度	.153 <sup>+</sup>	1.000		
③高等学校での「数学」の授業に対する印象(好きか嫌いか)	.164 <sup>+</sup>	.612**	1.000	
④高等学校での「数学」の授業に対する意識(得意不得意)	.244**	.554**	.833**	1.000

\*\* $p < .01$ , + $p < .10$ 

表11 情報の授業の満足度・好き嫌い・得意不得意と正答率の相関

	①	②	③	④
①リメディアルチェックテスト—正答率	1.000			
②高等学校での「情報」の授業の満足度	-.116	1.000		
③高等学校での「情報」の授業に対する印象(好きか嫌いか)	-.055	.617**	1.000	
④高等学校での「情報」の授業に対する意識(得意不得意)	-.088	.367**	.669**	1.000

\*\* $p < .01$

表12 ICT基礎スキルの習得の有無による正答率の違い

習った基礎的スキル	チェックあり		なし		t 検定結果
	平均正答率 (SD)	n	平均正答率 (SD)	n	
1. キー入力 (タイピング)	0.531 (0.212)	114	0.520 (0.218)	20	$t(132) = -0.23, p = .820, d = -0.055$
2. ワープロ (Word等)	0.529 (0.216)	93	0.530 (0.204)	41	$t(132) = 0.00, p = .999, d = 0.000$
3. 表計算 (Excel等)	0.560 (0.203)*	81	0.483 (0.218)*	53	$t(132) = -2.07, p = .040, d = -0.364$
4. プレゼンテーション (PowerPoint等)	0.568 (0.212)**	82	0.469 (0.199)**	52	$t(132) = -2.68, p = .008, d = -0.473$
5. Web作成	0.590 (0.200)	17	0.521 (0.213)	117	$t(132) = -1.26, p = .208, d = -0.326$
6. プログラミング	0.603 (0.188) +	20	0.517 (0.214) +	114	$t(132) = -1.68, p = .095, d = -0.406$

\*\* $p < .01$ , \* $p < .05$ , + $p < .1$ 

表13 パソコンを使い始めた時期による正答率の違い

習った基礎的スキル	高校入学以前		入学以後/使っていない		t 検定結果
	平均正答率 (SD)	n	平均正答率 (SD)	n	
パソコンの使用開始時期	0.548 (0.205)	93	0.487 (0.224)	41	$t(132) = 1.57, p = .120, d = 0.292$

仮説(6). コンピュータを使い始めた時期が早い学生はチェックテストの正答率が高い

コンピュータを使い始めた時期(表7)の回答に対して、学生を高校以前に使い始めた群と高校入学以後に使い始めた(もしくは使っていない)群の2群に分けて、それぞれの正答率を対応無しt検定によって比較した。その結果、正答率にはやや違いがみられたものの、有意な違いはみられなかった(表13)。

仮説(7). データサイエンスの各スキルを「身に付けている」と認識している学生はチェックテストの正答率が高い

本調査で身に付けているかどうかを聴取したDSの6種類の各スキル(図7)の4段階尺度の回答及び、6種類すべてのスキルにおける各回答の加算平均に対して、正答率との間の相関分析を行った。その結果(表14)、「データ・資料を収集する能力」と正答率、そして「グラフや表の数値を読み取る能力」と正答率との間に有意な弱い正の相関が確認された。これらのスキルを身に付けていると認識していることがDSに

関する数学的基礎能力の向上と関連のある可能性が考えられる。また、全スキルの加算平均においても有意傾向ではあるが弱い正の相関が確認できた。

仮説(8). 様々なDS用語を理解している学生はチェックテストの正答率が高い

DS用語の理解度チェック(図9)の回答に対して、各用語の理解度を

「1. 理解している」→1点

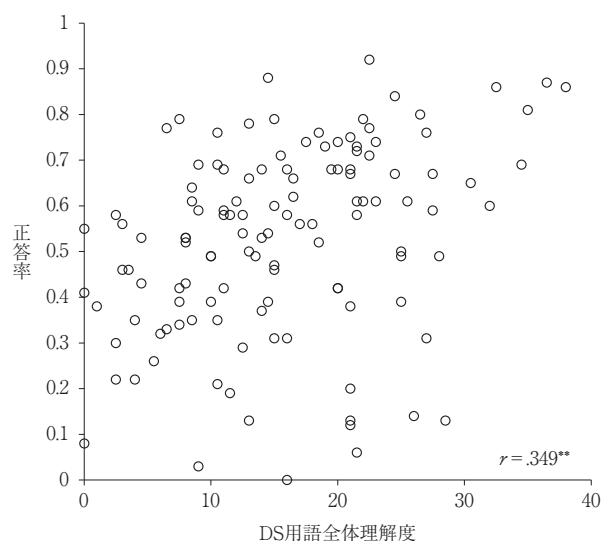
「2. ある程度理解している」→0.5点

「3. 何も分からない」→0点

と数値化して42語全ての用語の分を加算し、「DS用語全体理解度」と定義した。その値と正答率の相関分析を行ったところ、低い正の相関( $r = .349$ )が確認された。図10に散布図を示す。この結果より、DSに関連する用語を学ぶことがDSに関する数学的基礎能力の向上に繋がる可能性が期待できる。

表14 DSスキルの自己認識と正答率の相関

	正答率との相関係数
リメディアルチェックテスト正答率	1.000
データ・資料を収集する能力	.211*
グラフや表の数値を読み取る能力	.255**
データ収集のための実験や調査などの企画立案の能力	-.026
パソコンの表計算ソフト等を使い、簡単なデータ集計や分析をする能力	.065
要因分析や問題・課題解決の情報を抽出する能力	.078
分析結果を人に伝える能力(コミュニケーション・プレゼンテーション能力)	.054
全スキルの加算平均	.159+

\*\* $p < .01$ , \* $p < .05$ , + $p < .1$ 図10 DS用語全体理解度と正答率の散布図 ( $n = 125$ )

## 4. 考察

本章では、前章で得られた結果についてまとめ、考察を行う。

### 4.1 大学におけるDSリメディアルの実施方法

DSリメディアルチェックテストの結果、各設問の難易度と正答率の関係が明らかになった(図2)。この中で、小学3年～5年の難易度の問題で正答率が低かった問題(問題番号4～7)に関しては四捨五入等の学習者のケアレスミスによる誤りが多かった可能性が窺えることから、大学においてDSのリメディアル教育を行う際には、就職活動で役立つとされるSPIテストを活用することも踏まえて学生側にもメリットのある形で補習授業やオンライン学習を促すやり方により比較的短時間でカバーできる可能性が考えられる。

これより深刻なのは、図2における問題番号15の度数分布票と相対度数、問題番号17の確率計算、問題番号19の累積相対度数、そして問題番号20の度数分布表のほうである。これらは正答率が低い理由として、該当する単元を持つ数学Bなどの科目を履修していないことが考えられる。実際に「確率分布と統計的な推測」の単元を持つ数学Bの履修率は約4割と低い(図3)。従ってこれらの確率統計の内容をリメディアル教育として実施するためには、高校における数学科目が未履修であることを踏まえて教育内容を検討する必要がある、履修した学生と未履修の学生では別の方法を検討することも重要と考えられる。表8より数学Ⅱ、数学B、そしてDS関連の3つの単元(「データの分析」、「場合の数と確率」、「確率分布と統計的な推測」)を履修した学生の正答率は未履修の学生よりも高い結果が得られていることから、これらの内容を重点的に学べるような教材やオンラインコースを用意する手も有効と考えられる。今回、調査対象とした大学1年生の中で、約3割の学生が高校の数学が「嫌いだった」、「苦手だった」と回答したことから、いかに意欲を持たせてかつ適切な難易度のDSリメディアル教育を実施できるか、大学側に工夫が求められていると考えられる。小・中・高等学校におけるわずか42のDS関連用語(図9)においても難易度によらずその理解度には大きなばらつきがある状況からも、初等中等教育においてDS関連の内容は学習の取りこぼしが多いことが窺える。このような状況の中、大学側では学習者の習熟度に応じたりメディアルの実施が求められていると考えられる。

### 4.2 DSに関連する情報教育

今回、調査対象とした大学1年生のうち、約8割(99/134)の学生が教科「情報」を少なくとも1科目は履修していた(表9)。しかしその教育内容は表5で示したような専門性のある授業はあまり行われておらず、教科書に沿った座学中心の授業もしくは、表6

のようなキー入力、ワープロ、表計算、プレゼンテーションといった基礎的スキルの習得に留まっていた可能性が考えられる。しかしその一方でこれらの基礎的スキルを習得することは重要であり、特に「プレゼンテーション」、「表計算」、「プログラミング」の習得がDSに関する数学的基礎能力の向上と関連することが示された(表12)ことから、これらのスキルを習っていない学生に対して大学は入学後のなるべく早い時期に対策を打つ必要があると考えられる。教科「情報」やコンピュータは高校数学と比べると「嫌い」や「苦手」とする学生が少ないこと(図5、図6)から、数学のリメディアル教育よりもICTの基礎的スキルの習得は進めやすいと考えられる。しかし、大学でICTスキルの習得をリメディアルとして行うのであれば、基礎的操作のみならず「データ・資料を収集する能力」や「グラフや表の数値を読み取る能力」(表14)などDSに関する数学的基礎能力の向上と関連性の高いDSスキルの習得と繋げるような取り組みが有効と考えられる。

## 5. 結論

本論文では、データサイエンス教育における基礎学力の把握及び影響要因とニーズ調査のために、データサイエンス分野のリメディアルチェックテストとアンケートを開発し、2つの大学の1年生計140名を対象にテストと調査を実施した。その結果、DSに関連する高校数学のリメディアル分野において学生のつまづきやすい事項が示唆されると共に、DSに関連する高校数学の科目や単元の受講が一部に留まっている現状、そしてDSスキルへのニーズとDS関連用語の理解度の実情が明らかになった。さらに、テストの正答率と調査項目の関連性を調べた結果、数学Ⅱと数学B、そしてDSに関連する数学Ⅰの「データの分析」、数学Aの「場合の数と確率」、数学Bの「確率分布と統計的な推測」の単元を履修した学生、そしてプレゼンテーションソフトと表計算ソフトを学校で習った学生の正答率が有意に高く、これらの事項がDSに関する数学的基礎能力の向上と関連のある可能性が示唆された。

今後の課題としては、テストの正答率が高い学生の傾向を探ることでDSのリメディアル教育の在り方を明らかにできると考えられる。また、学習指導要領が改訂されたことを踏まえ、高校における教科「情報」等におけるDS分野の取り扱い方の調査分析も併せて行う必要があると考えられる。

## 謝辞

調査に協力して頂いた2大学の教職員の方々に深く感謝する。また、本研究は放送大学学習教育戦略研究所 研究課題「データサイエンスのモデルカリキュラム準拠リテラシーレベルの準備および補完学習としてのリメディアル教育のニーズや課題に関する調査」の

支援を受けた。

## 注

本論文は (10) および (11) の研究報告に対して研究内容の追加及び追記を行い書き直したものである。

## 参考文献

- (1) 放送大学：“初級レベルの数理・データサイエンス・AIについて学べるインターネット配信公開講座のご案内”、<https://www.ouj.ac.jp/hp/special/article/datascience.html> (2021) (参照日：2021年10月1日)
- (2) 日本数学会：“「大学生数学基本調査」に基づく数学教育への提言”、<https://mathsoc.jp/comm/kyoiku/chousa2011/>、(2012) (参照日：2021年10月1日)
- (3) 芳沢光雄：“「%」が分からない大学生～日本の数学教育の致命的欠陥～”、光文社 (2019)
- (4) 公益財団法人日本数学検定協会、実用数学技能検定(数学検定・算数検定)、<https://www.su-gaku.net/>、(2021) (参照日：2021年10月1日)
- (5) 一般財団法人統計質保証推進協会、統計検定、<https://www.toukei-kentei.jp/about/>、(2021) (参照日：2021年10月1日)
- (6) 古川徹：“文系学生の統計学の教育の問題点”、山陽論叢、23巻97-112 (2017)
- (7) 河村一樹、情報分野における高大接続のためのプレースメントテストの実施と評価、東京国際大学人間科学・複合領域研究第4号、(2019)
- (8) 竹内光悦・末永勝征、データサイエンス教育に関する調査結果からみる統計基礎教育の現状、統計数理、第66巻第1号、pp. 107-120、(2018)
- (9) 木村信行・伊澤毅、新入生に対する数学用語理解度調査、日本工学教育協会 平成28年度工学教育研究講演会講演論文集講演番号2F15、(2016)
- (10) 辻靖彦、秋葉広人、芝崎順司、松野由希、村上玲、矢尾板俊平、文系大学生を対象としたデータサイエンス教育のためのリメディアルチェックテストの開発と実施、教育システム情報学会第46回講演論文集P1-14、pp. 27-28、(2021)
- (11) 辻靖彦、秋葉広人、芝崎順司、複数機関の大学生を対象としたデータサイエンスに関する数学的基礎能力の測定とその影響要因の調査、日本教育工学会2022年春季全国大会、3-5101-5、(2022)

(2021年11月12日受理)