

姫路西高等学校におけるデータサイエンス教育の学習成果の検証 ～大学共通テスト『情報』の観点から～

林宏樹^{1,2} 渡辺博芳²

概要：令和7年度大学入学選抜に係る大学入学共通テストにおいて、必修科目「情報Ⅰ」の内容が『情報』として出題される。兵庫県立姫路西高等学校（以下、姫路西）は、スーパーサイエンスハイスクール事業において、データサイエンス教育を実践し、データサイエンスを基盤とした課題研究を行っている。本研究は、大学入学共通テスト『情報』を解く力の観点で、姫路西の取組における学習成果の一つの側面を評価した。（独）大学入試センターが令和3年3月24日に発表したサンプル問題の類似問題を作成し、7月にサンプル問題、12月に独自の類似問題による試験を実施した。試験の結果を分析し、姫路西の取組により、大学入学共通テスト『情報』の観点での学習成果は十分にあったことが示された。

キーワード：データサイエンス教育、大学共通テスト、情報、データの活用

A Validation of learning outcomes of Data Science Education in Himeji Nishi High School In teams of the University Entrance Examinations of “Information”

HIROKI HAYASHI^{†1} HIROYOSHI WATANABE^{†2}

1. はじめに

平成30年3月に新しい高等学校学習指導要領（以下、「新学習指導要領」）が告示され、高等学校等においては、令和4年度から新学習指導要領が年次進行で実施される。大学入学共通テストは、大学への入学志願者を対象に、高等学校等の段階における基礎的な学習の達成の程度を判定し、大学教育を受けるために必要な能力について把握することを目的として実施するものであることから、令和7年度大学入学共通テストから新学習指導要領に対応したものとする必要があり[1]。

教科「情報」における新学習指導要領では、2科目（「情報Ⅰ」、「情報Ⅱ」）が設定され、これらのうち「情報Ⅰ」が必修科目とされている[2]。

大学入学共通テストにおいて、国語、数学、英語のような基礎的な科目として必修科目「情報Ⅰ」（コンピュータの仕組み、プログラミング等）を追加するとされ、必修科目「情報Ⅰ」の内容が『情報』として出題される[1]。

このような中で、令和2年度、文部科学省によって学校全体でデータサイエンス教育に取り組むカリキュラム設計のモデル校として、兵庫県立姫路西高等学校（以下、姫路西）が採択された。これを受けて、第一著者はデータサイエンス（以下、DS）のカリキュラム開発や授業実践に取り組んでいる[3]~[5]。

本稿では、姫路西におけるDS教育を受けた生徒の学習

成果の一つの側面として、筆記試験によって測定できる力の観点から評価することを目的とする。具体的には、（独）大学入試センターが令和3年3月24日に発表したサンプル問題『情報』の第3問（以下、サンプル問題）と、独自で作成した類似問題を用いて評価を行った。

2. 姫路西におけるDS教育

2.1 DS教育におけるカリキュラム

姫路西では、情報、数学、総合的な探究の時間によるカリキュラムマネジメントによってDS教育を行っている。第1学年ではデータサイエンスの基礎・基本を身につけ、第2学年ではデータサイエンスを基盤とした課題研究を実践している。

DS教育を実践する上で、『高等学校学習指導要領 解説 理数編』における「データ分析」に書かれている統計的探究プロセス[6]の5つの段階からなる「問題（Problem）－計画（Plan）－データ（Data）－分析（Analysis）－結論（Conclusion）」に基づき、DS教育を体系化している。プロセスの授業内容には、令和4年度入学生から実施される「情報Ⅰ」の「ネットワークとデータの活用」、「情報Ⅱ」の「情報とデータサイエンス」の内容に加え、「数学Ⅰ」のデータの分析、「数学Ⅱ」の統計的な推測の内容をDS教育の視点で効果的な順序に構成した授業実践を提案する³⁾。2年間におけるDS教育の年間計画を表1、表2に示す。

¹ 兵庫県立姫路西高等学校
Hyogo Prefectural Himeji Nishi Senior High School.

² 帝京大学 Teikyo University

表1 第1学年のカリキュラムと従来科目の対応表

時期	内容	従来科目
4月	データの分析	数学I
5月	MS-Excelの知識・技能	社会と情報
6月	相関係数・標準偏差の活用 単回帰分析法の理解	数学I 情報I
7月	探究活動・ポスター発表	総合的な探究の時間
9月	情報モラル・研究モラル	社会と情報
10月	データを活用した問題解決探究	
11月	スライド発表	総合的な探究の時間
12月	発表評価アンケート解析	社会と情報
1月	PPDACサイクルの復習	総合的な探究の時間
2月	新価値創造講座	
3月		

表2 第2学年のカリキュラムと従来科目の対応表

時期	内容	従来科目
4月	データサイエンスの復習	社会と情報
5月	回帰分析法の理解	情報I
6月	標準化の理解 統計的仮説検定の理解 クラスタリングの理解 主成分分析法の理解	数学B 情報II
7月	探究活動 ポスター発表	総合的な探究の時間
9月	Python活用によるデータ処理	社会と情報
10月	統計的な推測	総合的な探究の時間
11月	データ活用した探究活動	
12月	ポスター発表会	総合的な探究の時間
1月	発表評価アンケート解析	社会と情報
2月		
3月		

本教育実践は2020年度に入学した生徒に対するものなので、従来科目の教科情報については基本的には社会と情報と対応づけているが、情報Iと情報IIの内容も取り入れているため、表1と表2では混在して記述している。

2.2 学習成果の評価の方針

評価は、「知識及び技能」、「思考力、判断力、表現力」、「学びに向かう力、人間性等」の観点に基づく評価を行う。姫路西のDS教育では、「知識及び技能」は筆記試験、「思考力、判断力、表現力」はデータサイエンスを基盤とした探究活動による研究ポスターのルーブリック評価、「学びに向かう力、人間性等」は日常の探究活動の行動観察や研究記録簿によって評価を実施する。

「知識及び技能」の筆記試験は、2.1の年間計画に示した

通り、数学Iの「データの分析」、数学Bの「統計的な推測」、情報Iの「データの活用」、情報IIの「情報とデータサイエンス」に関する筆記試験やエクセルを活用した実技試験によって、総合的に評価する。

本稿では、「知識及び技能」を評価するための筆記試験の一部の実施内容について述べる。

3. 大学共通テストサンプル問題の類似問題の作成

3.1 サンプル問題『情報』第3問の内容

大学入試センターで発表されたサンプル問題『情報』[1]の第3問は次のような問題から構成されている。

問1は、多くの項目があるデータを可視化した複数の散布図や相関係数から項目間の関係などを考えさせる問題である。

- ・細問ア・イ・ウは、相関係数を読み取る問題である。
- ・細問エは、多くの項目があるデータを可視化された図の読み取りからの正誤問題である。

問2は、回帰直線からデータの関係性や予測値、残差について考えさせる問題である。

- ・細問オカは、回帰直線からグループの傾向の違いを読み取らせる問題である。
- ・細問キは、回帰直線から予測値の差を求める問題である。
- ・細問クケは、実際の値と予測値との差である残差を求める問題である。

問3は、基本統計量の読み取りから判断する問題である。

- ・細問コ・サは、四分位数や標準偏差などの基本統計量から、データに含まれる傾向を読み取らせる問題である。

問4は、四分位数を基にしたデータの散らばりから傾向を考えさせる問題である。

- ・細問シは、クロス集計表の読み取りによる正誤問題である。
- ・細問ス・セソは、クロス集計表からのデータの傾向を読み取らせる問題である。

3.2 作成した問題の内容

サンプル問題第3問に類似した問題（以下、独自問題）を作成した。独自問題のうち、6つの細問はサンプル問題とほぼ同様な問題である。サンプル問題の細問と独自問題の細問の対応表を表3に示す。サンプル問題のシ〜ソに対応する問題は作成できず、その代替として、サンプル問題との全体的な難易度で同じような力を測り、力の伸長を調べるために独自問題のエ・タ・チを設定した。

表3 サンプル問題と独自問題の対応表

対応番号	サンプル問題	独自問題
1	ア・イ	ア・イ
2	ウ	ウ
-	エ	
-		エ
3	オカ	オカキ
4	キ	クケコ
5	クケ	サシス
6	コ・サ	セ・ソ
-	シ・ス・セソ	
-		タ・チ

3.2.1 対応番号1・2の内容

サンプル問題の問1に対応する複数の散布図や相関係数に関するデータは、2010年～2019年の各月（1月～12月）における日本の各月の気温（日中最高気温の平均、単位は℃）、デザート（ケーキ・チョコレート・アイスクリーム・ようかん）の平均購入金額（各都道府県の家計調査の月ごとの1日当たりの平均購入金額、単位は円/日）を使用した（図1）⁵⁾。

- ・独自問題のア・イは、相関係数に基づいて相関関係を判断させ、2番目に正の相関が強い組合せを判断する問題である。
- ・独自問題のウは、最も負の相関が強い関係を判断する問題である。

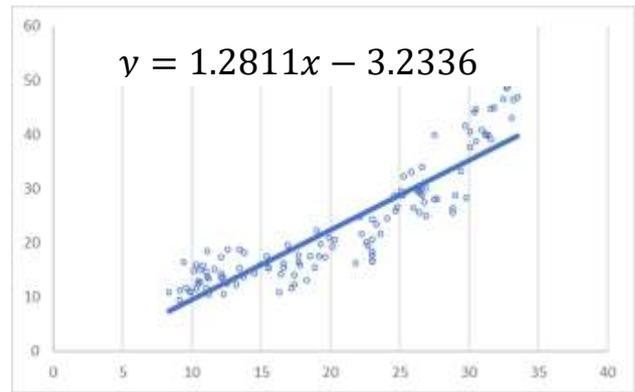
3.2.2 対応番号3・4・5の内容

サンプル問題の問2に対応するデータは、図2の気温とアイスクリームの平均購入金額のデータの関係性から、予測値、残差について考えさせる問題とした。

- ・独自問題のオカキは、気温が10℃高くなるごとに、その月のアイスクリームの増加する平均購入金額を予測する問題である。
- ・独自問題のクケコは、アイスクリームの平均購入金額が30円するとき、回帰直線から予測できるその月の気温を求める問題である。
- ・独自問題のサシスは、気温が20℃のときのアイスクリームの平均購入金額が25円の実際の値と、予測値との差である残差を求める問題である。

3.2.3 対応番号6の内容

サンプル問題の問3に対応する問題は、表4の分析シートに示す基本統計量から読み取れることを考えさせる問題である。散らばりを四分位範囲や標準偏差の視点で判断したり、第1四分位数・第2四分位数、中央値、平均値をそれぞれ比較する力を問う。



横軸：気温（℃）
 縦軸：アイスクリームの平均購入金額（円）

図2 気温とアイスクリームの平均購入金額の回帰直線

表4 2010年～2019年のデザートの平均購入金額に関する基本的な統計量（分析シート）

	ケーキ	チョコレート	アイスクリーム	ようかん
合計	2252.49	1808.35	2787.4	239.2
最小値	12.39	4.32	9.45	0.82
第1四分位数	14.9	8.92	14.94	1.36
中央値	16.79	12.79	19.385	1.67
第3四分位数	18.03	17.05	29.31	2.56
最大値	44.84	50.18	48.81	4.23
分散	61.38	105.18	114.81	0.75
標準偏差	3.133	8.125	14.368	1.203
平均値	18.77	15.07	23.23	1.99

3.2.4 その他の問題

サンプル問題のエに対応する問題として、独自問題のタを作成した。サンプル問題では、サッカーのワールドカップに関する決勝進出チームと予選敗退チームのグループごとに分かれたデータを使用している。そのため、独自問題における平均購入金額を5月～10月（I期）と11月～4月（II期）に分けた散布図（図3）から、I期とII期を比較する問題を作成した。

しかし、サンプル問題のエと独自問題のタでは解答する順番が大きく異なることと、選択肢の難易度が異なると判断したため、本稿では対応させなかった。具体的には、サンプル問題のエでは、選択肢④が、「それぞれの散布図の中で、決勝進出チームは黒い四角形（■）、予選敗退チームは白い円（○）で表されている。」とあり、独自問題のタでは、選択肢④が、「平均購入金額と年間の気温の相関だけで判断せず、I期・II期と時期を細かく分けた相関で判断すると、強い相関がある時期がみられるデザートがある。」と設定した。サンプル問題では層別に分けた読み取りだけを求めているが、独自問題では層別に分けた読み取りに加え、データの読み取りを判断させる問題となり、独自問題の方が難易度は高かったと判断した。

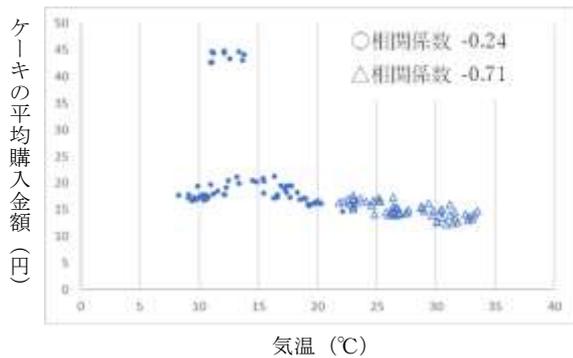


図3 気温とケーキの平均購入金額の関係

4. テストの実施結果

4.1 テストの実施方法

テストは定期考査で2回行った。1回目が令和3年7月5日、2回目が令和3年12月6日である。対象者は第2学年の普通科238名である。1回目のテストではサンプル問題、2回目のテストでは独自に作成した問題を解かせた。

また、テスト前の活動として、1回目の事前授業では単回帰分析の授業を50分×2コマ行っており、2回目のテストの事前授業では1回目の事前授業の資料を見ておくように指示するだけに留めた。

最後に、2回目のテスト終了後、受験者238名のうちの77名に各対応問題のサンプル問題と独自問題の難易度について意識調査を行った。難易度の回答は「簡単であった」、「普通」、「難しかった」を選択させた。

4.2 テストの結果

4.2.1 1回目の受験結果

対応していない問題を含めた全体正答率は58%であった。各問題の細問ごとの正答率を図4に示す。なお、「アイ」と解答する場合や「ア」と「イ」の両方を完全解答して正解となる場合は「ア」と表示している。対応がとれる6問に焦点をおいた全体の正答率は62%であった。対応がとれる問題は、解答ア・イ、ウ、オカ、キ、クケ、コ・サである。

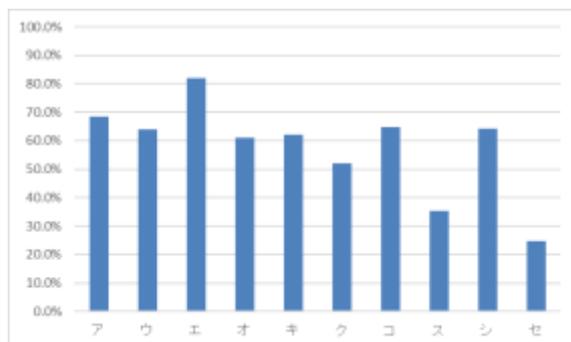


図4 1回目の各問の正答率

4.2.2 2回目の受験結果

対応していない問題を含めた全体正答率は60%であった。各問題の細問ごとの正答率を図5に示す。対応がとれる6問に焦点をおいた全体の正答率は70%であった。対応がとれる問題は、解答ア・イ、ウ、オカキ、クケコ、サシ、セ・ソである。

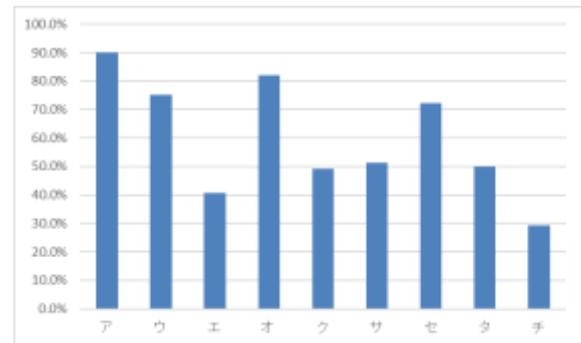


図5 2回目の各問の正答率

4.2.3 1回目と2回目の全体の比較

(1) 対応していない問題を含めた全体正答率

解答数は1回目が10問、2回目が9問であり、正答率によって1回目と2回目を比較する。対応していない問題を含めた全体正答率の平均は、1回目58%、2回目60%であり、若干増加した。

個々の生徒を見ると1回目から2回目で正答率が向上した生徒は44%、変化しなかった生徒は1%(2回とも満点)、低下した生徒は55%であり、低下した生徒が多い。

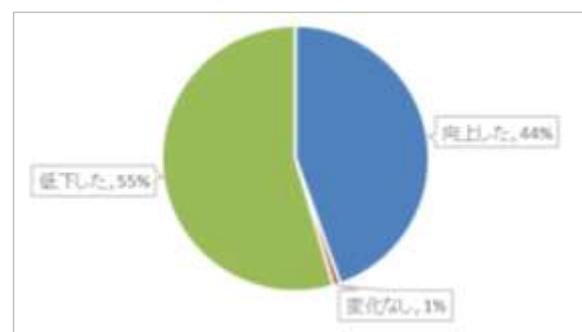


図6 1回目から2回目の生徒の変容

一方で、得点分布を箱ひげ図(図7)で示すと、1回目に比べ、2回目は得点分布の散らばりが小さくなっている。平均点は若干向上しているが、中央値が下がっていることから、上位層は得点を上げたが下位層の人数が増えたことが示唆される。

正答率

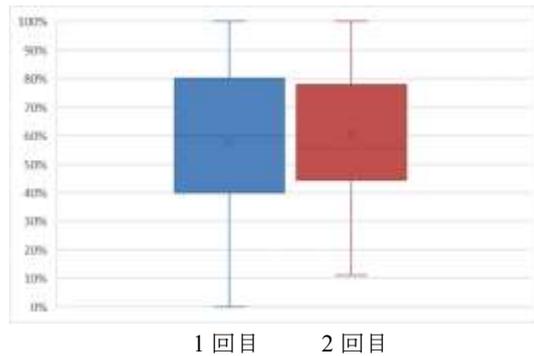


図7 1回目と2回目の得点の箱ひげ図

(2) 対応問題に焦点をおいた全体結果

対応がとれる問題は全部で6問である。対応した6問の全体正答率の平均値は、1回目62%、2回目70%であり、有意水準5%としてt検定を行うと、統計的に有意差($p=0.590 \times 10^{-4}$)があったため、向上したと判断できる。

また、1回目から2回目で正答率が向上した生徒は47%、変化しなかった生徒は21%、低下した生徒は32%であった(図8)。概ね向上した生徒が多いことが読み取れる。

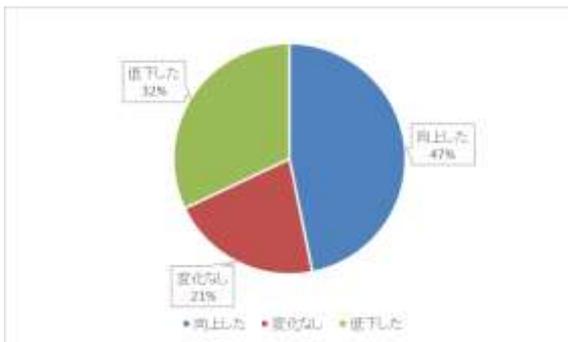


図8 1回目から2回目の生徒の変容

得点分布を箱ひげ図(図9)で示すと、得点の散らばりの違いは見受けられない。中央値は変化せず、平均値が向上したため、全体としては向上したと判断できる。

正答率

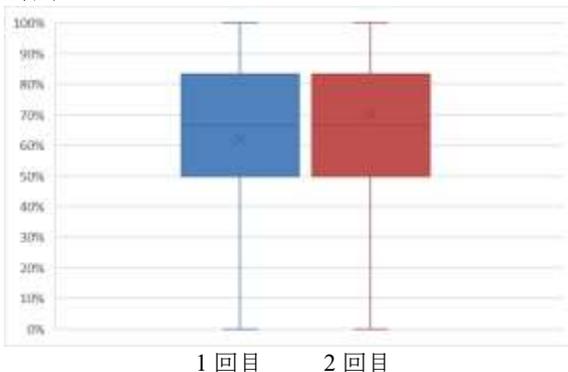


図9 1回目と2回目の得点の箱ひげ図

4.2.4 対応する個別問題の結果

細問ごとの正答率を図8に示す。正答率の平均は、対応番号1・2・3・6は向上し、対応番号4は下がり、対応番号5は変化がなかった。

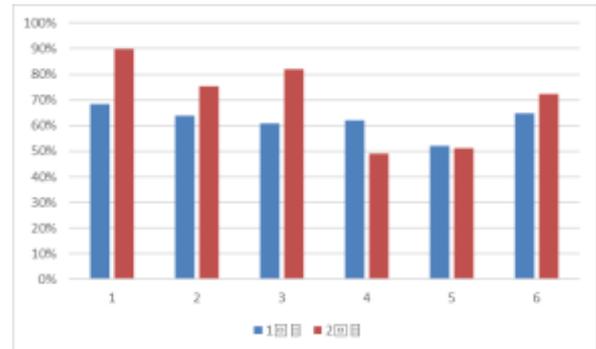


図10 1回目と2回目の各対応問題の正答率

次に、対応した問題の正誤の変容を表5と図11に示す。ここで分類は次の通りである。

- A : 1回目・2回目ともに正解
- B : 1回目不正解・2回目正解
- C : 1回目正解・2回目不正解
- D : 1回目・2回目ともに不正解

表5 1回目と2回目の個別問題の変容(割合)

対応番号	A	B	C	D
1	64%	26%	5%	5%
2	50%	25%	14%	11%
3	51%	31%	10%	8%
4	33%	16%	29%	21%
5	28%	24%	24%	24%
6	49%	24%	16%	12%

割合は小数第一位で四捨五入したため、各問題の割合の合計が100%になっていない場合がある。

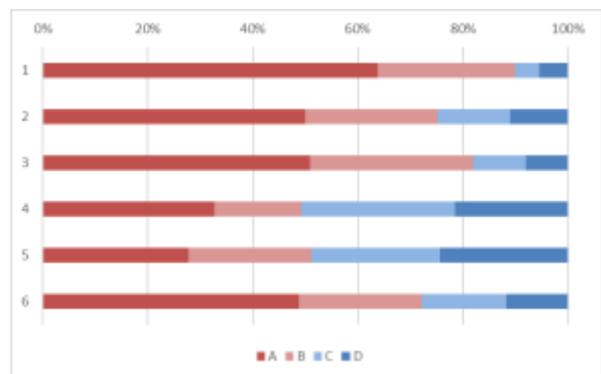


図11 1回目と2回目の個別問題の変容

結果から対応した問題 4・5 以外は A・B の生徒が多いことがわかる。対応した問題 4・5 に関しては、A・B と C・D の割合がほぼ同じであり、4 においては、B より C の方が多いことがわかる。

4.3 アンケートによる意識調査結果

図 12 にアンケート結果を示す。なお、問題 1-1 とは、対応番号 1 のサンプル問題（1 回目）を表し、問題 3-2 とは対応番号 3 の独自問題（2 回目）を表す。

全体では、受験者の意識として、対応番号ごとに 1 回目、2 回目の「難しかった」の回答を比較すると大きな差はないため、1 回目、2 回目の難易度はほぼ変わらなかったと判断できる。

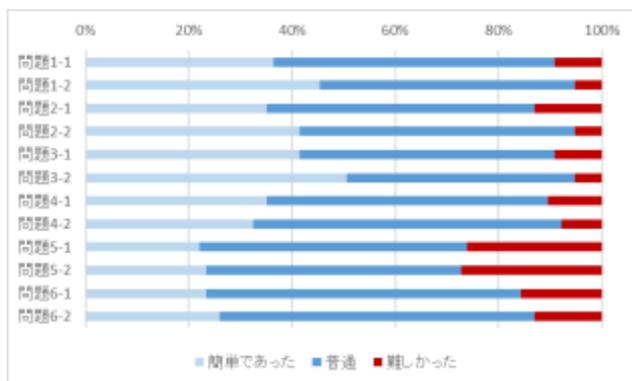


図 12 意識調査の結果

5. 考察

5.1 大学共通テストの観点での学習成果

対応していない問題を含めた全体正答率の結果は、1 回目、2 回目ともにある程度よい成果であったと判断している。1 回目（サンプル問題）については、令和 3 年度大学入学共通テストの数学 I・A の平均点が 57.68 点であること、本校の生徒が第 3 学年の受験勉強で最終的に力を向上させていることを考慮すると、第 2 学年 7 月時点で教科は異なるものの共通テストの平均点と同程度の平均正答率を達成しているからである。また、2 回目（独自問題）についても、共通テストサンプル問題と同程度の難易度であることを意識して作成した問題に対して、第 2 学年 12 月時点で平均正答率 60% であることは、同様に考えると、ある程度よい成果であると判断できる。

対応問題に焦点をおいた全体正答率を比較する。4.3 から受験者の意識は、「簡単であった」「普通」の回答の割合から、サンプル問題と独自問題で対応する問題の難易度はほぼ等しいと判断した。

対応問題に焦点をおいた全体正答率を比較すると、中央値は変化していないが、平均値は向上したので、7 月から 12 月に行われた探究活動においても、教育成果は上がった

と判断できる。図 9 からは得点分布の散らばりは見受けられないが、標準偏差は 1 回目が 1.8、2 回目が 1.3 と小さくなり、2 回目の方が集団として向上した。これは、DS を基盤とした課題研究を通じてデータの可視化、整理整形、解析の過程を経験する上で培った成果であると考えている。

また、各テストの事前準備として、1 回目は事前に理論的内容を学んだ直後であったが、2 回目は理論を活用する課題研究を実施し、7 月の内容を復習するように指示はしていたが、授業時間として復習の時間を設けていなかった。このことから、DS を基盤とした探究活動を実施することにより、特別に筆記試験のための学習をしなくても筆記試験の平均値が変化せず得点分布の散らばりが小さくなったことは、本教育実践の成果の 1 つであると考えられる。

以上のことから、姫路西での DS 教育の取組は、大学共通テストの観点での学習成果は十分にあったと判断した。

5.2 誤答の生徒が多い問題

4.2.4 の結果から、C の生徒が B の生徒より多い対応番号 4 と、B の生徒と C の生徒の割合が等しい対応番号 5 について分析する。なお、問題の難易度を判断する 1 つの基準とした 4.3 意識調査から、対応番号 5-1、5-2 はともに、他の問題と比べ、「難しかった」と選択した生徒は 5-1 が 26%、5-2 が 27% と割合が高かった。各対応問題の「難しかった」を選択した生徒の人数は 5-1 が 20 人、5-2 が 21 人であり、平均 9.41 人に対して平均+2σ が 20.4 人であることから、他の問いに比べ、明らかに「難しかった」と回答した生徒が多かったと判断できる。

(1) 対応番号 4 の誤答分析

C の生徒が B の生徒より多い対応番号 4 に該当する独自問題クケコの誤答を調査すると、不正解者のうち 36% が計算ミスであり、理解不足である生徒は 64% であった。

この細問は正答率 49% であり、不正解者は、121 名であった。回帰式 $y=1.2811x-3.2336$ をどのように扱っているのか考察する。正答は 25.9 であり、y に 30 を代入し、x を求めればよい。

第一に、正答の 25.9 を除く 25.0~26.2 の解答者は 33 名であった。各係数が小数第 4 位まで示した式であるため、計算ミスや四捨五入のミスの可能性であると考えられる。

第二に、20.8~21.0 の解答者は 18 名であった。y に 30 を代入し、3.2336 を引き、1.2811 で割ると、20.9 となる。つまり、3.2336 を不足処理を引いた計算過程の 3.2336 の移行ミスである。

第三に、35.2~36.2 の解答者は 11 名であった。x に 30 を代入すると、y の値が 35.2 となる。与えられた条件が説明変数であるのか、目的変数であるのか理解していない誤答である。

つまり、不正解者の 141 名のうち、第一、第二の誤答者の 51 名は計算ミス、残り 90 名は理解不足が原因であるこ

とがわかった。

(2) 対応番号5の誤答分析

Bの生徒とCの生徒の割合が等しい対応番号5に該当する独自問題サシスの誤答を調査すると、不正解者のうち13%が計算ミスであり、残り87%が理解不足であった。

この細問は正答率51%であり、不正解者は、116名であった。正答は2.61であり、 x に20を代入し、 y の値を求め、その値と25との差を求めればよい。

第一に、0.62の回答者は28名であった。これは、 x に20を代入し切片の-3.2336を処理しない値と実測値25との差である。回帰式から予測値を求める理解ができていない誤答である。

第二に、3.61の回答者は9名、2.61を除く2.60~2.66が6名である。これは、単純な計算ミスであると考えられる。

つまり、不正解者116名のうち、計算ミスは15名、理解不足は101名であるとわかった。

(3) 誤答分析からみる今後の指導改善

理解不足での誤答者は、対応番号4が30%、対応番号5が42%である。これらの理解不足に対応する方策の一つは、回帰直線、予測値などの用語や、回帰直線における説明変数や目的変数の理解に関する学習時間を増加させることである。他の方策としては、探究活動において各生徒が必ずしも回帰分析を活用しているわけではないが、回帰分析を題材とした探究の演習を充実させることが考えられる。間接的ではあるが、これにより第2学年において、回帰分析を用いた探究を行う生徒の探究事例が増えることで、発表会などを通じて他の生徒に波及し、筆記試験に関わる知識の定着率の増加に寄与することが期待される。

6. おわりに

本稿では、姫路西における2年間のDS教育の学習成果の1つとして、大学共通テストの観点からの学習成果の検証を目的として、大学入試共通テストのサンプルと、独自に作成した類似問題の受験結果について分析を行った。その結果、本DS教育の取組は、大学共通テストの観点での学習成果は十分にあったと判断した。また、第2学年の探究活動の前後に実施したテストで対応した細問の正答率がやや向上していることから、探究活動においても大学共通テストの観点での学習成果が上がっていると思われる。

今後の課題として、探究活動における研究ポスターのルーブリック評価の分析を行い、本教育実践の成果を総合的に検証したい。また、回帰分析を用いた探究事例を主とした授業内容の改善を行う必要がある。

参考文献

- [1] 大学入試センター『平成30年告示高等学校学習指導要領に対応した令和7年度大学共通テストからの出題教科・科目について』(2021-03-24).
- [2] 文部科学省、『高等学校学習指導要領 解説 情報編』, 2018.
- [3] 林宏樹, 笹嶋宗彦, 大里隆也, 高等学校におけるデータサイエンス教育のカリキュラム開発と実践. 情報教育シンポジウム論文集, 2020, p. 146-150.
- [4] 林宏樹, 高等学校におけるデータサイエンスを基盤とした問題解決実践-AI時代を切り拓くための創造力の育成を目指す-, 情報処理, 2021, vol.62, no.10, p. 560-565.
- [5] 林宏樹, 学校設定科目「DS探究」「DS研究」におけるデータサイエンス教育の成果と今後の展望, 統計数理研究所共同研究レポート448, 2021, 統計教育実践研究第13巻
- [6] 文部科学省、『高等学校学習指導要領 解説 理数編』, 2018.
- [7] 笹嶋宗彦, 『データサイエンス入門』, 朝倉書店

付録

大学共通テストサンプル問題（第3問）に基づいて作成した独自問題

次の文章を読み、後の問い(問1～3)に答えよ。

将来経営者を目指すデザート好きのH高等学校の田中さんは、「デザートのお店を経営するときに、どのような売り方をすれば売れるお店になるのか」というテーマについて研究している。田中さんは、1年間でどの時期にどのようなデザートが売れるのかデータに基づいて分析することにした。このデータは、2010年～2019年の各月（1月～12月）における日本の各月の気温（日中最高気温の平均、単位は℃）、デザート（ケーキ・チョコレート・アイスクリーム・ようかん）の平均購入金額（各都道府県の家計調査の月ごとの1日当たりの平均購入金額、単位は円/日）であった。田中さんは、このデータを基に、各項目間の関係を調べることにした。データ加工には、表計算ソフトウェアを活用し、表1のデータシートを作成した。

表1 2010年～2019年の各月におけるデータの一部
(データシート)

年	月	気温	ケーキ	チョコレート	アイスクリーム	ようかん
2010	1	11	19.68	14.71	11.68	1.77
2010	2	9.9	19.46	40.07	10.89	1.36
2010	3	13.2	21.19	13.19	12.35	1.13
2010	4	16.6	17.4	9.57	15.47	1.23
2010	5	23	17.77	7.52	24.26	2.03
2010	6	27.5	14.63	7.33	28.03	2.13
〰〰						
2019	9	29.4	16.3	12.47	33.2	1.53
2019	10	23.3	16.39	15.42	23.35	1.1
2019	11	17.7	19.6	18.37	17.7	1.33
2019	12	12.6	43.35	21.84	18.84	2.13

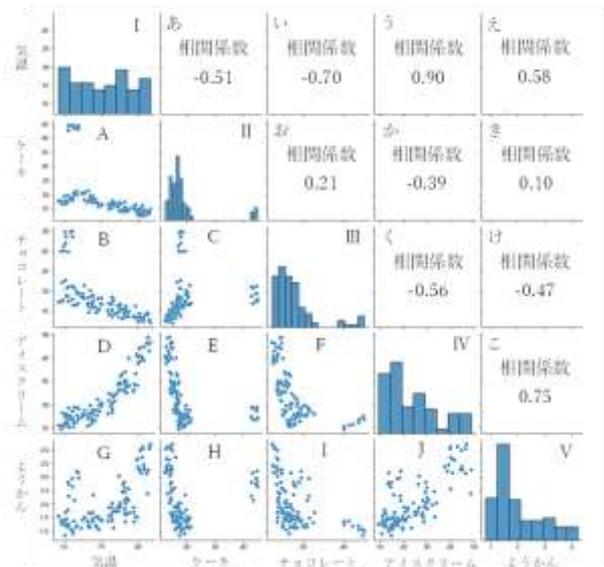


図1 項目間の関係

図1のI～Vは、それぞれの項目のヒストグラムである。【あ】～【こ】は、それぞれの二つの項目の相関係数である。またA～Jは、それぞれの二つの項目の散布図を描いている。例えば、図1のAは横軸を「ある月の気温」、縦軸を「その月のケーキの平均購入金額」とした散布図であり、それに対応した相関係数は【あ】で表している。

問1 次の文章の空欄 ～ に当てはまる最も適当なものを、後の解答群のうちから一つずつ選べ。ただし、 と の解答の順序は問わない。

相関係数に基づいて相関関係を判断すると、2番目に正の相関が強い組合せは、 と である。また、最も負の相関が強い関係を表している散布図は である。一般的に『 は相関がある』といえる。

・ の解答群

- ① 気温 ② ケーキ ③ チョコレート ④ アイスクリーム ⑤ ようかん

の解答群

- ① A ② B ③ C ④ D ⑤ E ⑥ F ⑦ G ⑧ H ⑨ I ⑩ J

エの解答群

- ① デザート a とデザート b は同時購入される
- ② デザート a とデザート b の売れる時間帯が同じ時間帯である
- ③ デザート a の平均購入金額が高い月はデザート b の平均購入金額も高い
- ④ デザート a の平均購入金額が、デザート b の平均購入金額に強く影響する
- ⑤ デザート a の平均購入金額が高いことがデザート b の平均購入金額が高くなる要因となる

問 2 次の文章の空欄「オカ」～「シス」に当てはまる数字をマークせよ。

田中さんは、図 1 から気温とアイスクリームの関係に着目し、さらに詳しく調べるために、アイスクリームの平均購入金額をその月の平均気温で予測する回帰直線を図 2 のように作成した。

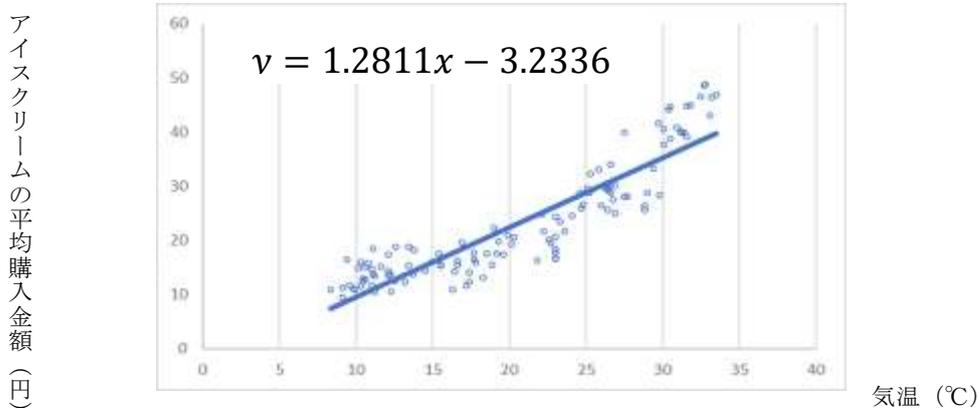


図 2 気温とアイスクリームの平均購入金額の回帰直線

この結果から、気温が 10°C 高くなるごとに、その月のアイスクリームの平均購入金額は、小数第 2 位を四捨五入すると約「オカ」「キ」円ずつ増加すると予測できる。

また、アイスクリームの平均購入金額が 30 円するとき、回帰直線から予測できるその月の気温は、小数第 2 位を四捨五入すると約「クケ」「コ」°C であることがわかる。

ある月の気温が 20°C のとき、アイスクリームの平均購入金額は 25 円であった。このとき、気温と回帰直線による予測値との差は、小数第 3 位を四捨五入すると、約「サ」「シス」円である。

問 3 次の文章の空欄「セ」・「ソ」に当てはまる最も適当なものを、後の解答群のうちから一つずつ選べ。ただし、解答の順序は問わない。

田中さんは、さらに分析を進めるために、データシートを基に、平均値や四分位数などの基本的な統計量を算出し、表 2 を作成した。

表2 2010年～2019年のデザートのパッケージの平均購入金額に関する基本的な統計量（分析シート）

	ケーキ	チョコレート	アイスクリーム	ようかん
合計	2252.49	1808.35	2787.4	239.2
最小値	12.39	4.32	9.45	0.82
第1四分位数	14.9	8.92	14.94	1.36
中央値	16.79	12.79	19.385	1.67
第3四分位数	18.03	17.05	29.31	2.56
最大値	44.84	50.18	48.81	4.23
分散	61.38	105.18	114.81	0.75
標準偏差	3.133	8.125	14.368	1.203
平均値	18.77	15.07	23.23	1.99

田中さんは、表2から「セ」と「ソ」について正しいことを確認した。

「セ」・「ソ」の解答群

- ① デザートの平均購入金額の散らばりを四分位範囲の視点で見ると、チョコレートよりケーキの方が大きいこと
- ② デザートの平均購入金額は、アイスクリームとようかんともに第2四分位数より平均値の方が小さいこと
- ③ デザートの平均購入金額をみると、アイスクリームはケーキの中央値より大きいこと
- ④ デザートの平均購入金額を標準偏差で比べると、チョコレートがようかんよりも散らばりが小さいこと
- ⑤ ようかんの平均購入金額の第1四分位数は、ケーキの中央値より小さいこと

問4 次の文章を読み、後の問い(a・b)に答えよ。

次に、気温が平均購入金額に関係していると考え、5月～10月（Ⅰ期）と11月～4月（Ⅱ期）にデータを分類して分析した。

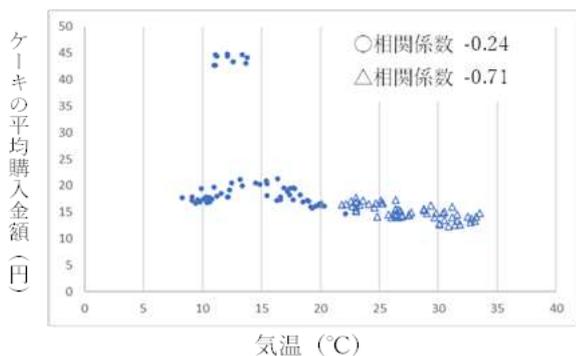


図3 気温とケーキの散布図

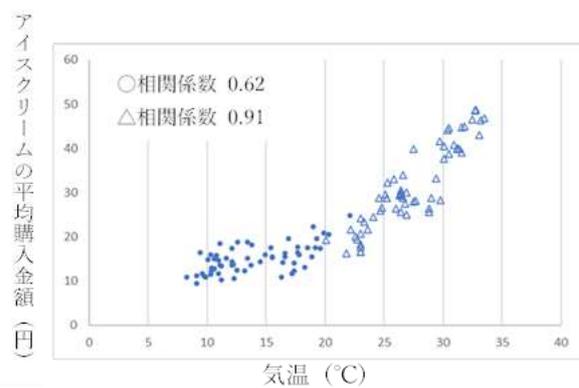


図4 気温とアイスクリームの散布図

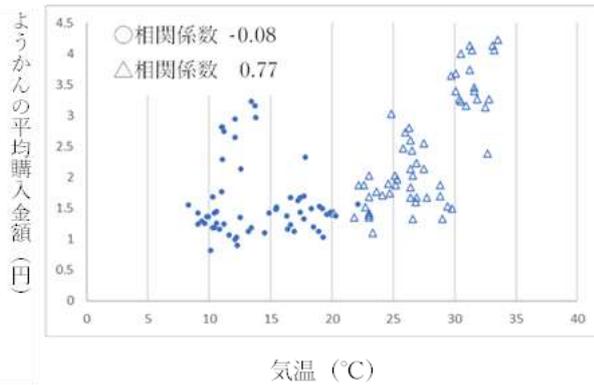


図5 気温とようかんの散布図

a. 図1, および図3～図5から読み取れることとして, 誤っているものを解答群から一つ選べ。

タの解答群

- ① 平均購入金額と年間の気温の相関だけで判断せず, I期・II期と時期を細かく分けた相関で判断すると, 強い相関がある時期がみられるデザートがある。
- ② ようかんの平均購入金額のI期は, 気温と相関があると判断できる。
- ③ 気温とアイスクリームにおける相関は, I期とII期をみると, I期の方が相関は強いと判断できる。
- ④ I期とII期を比較すると, ケーキ・アイスクリーム・ようかんともに, 気温と平均購入金額の相関においては, II期の方が相関は強いと判断できる。

b. 次の文の空欄「チ」に当てはまる最も適当なものを後の解答群から1つ選べ。

ケーキの平均購入金額が異常に高い群があり, それらを除いた場合, 気温との相関は「チ」。

チの解答群

- ① より強くなると判断できる
- ② より弱くなると判断できる
- ③ 正負が反転すると判断できる
- ④ ほとんど変わらないと判断できる
- ⑤ これだけの情報からは判断できない

独自問題の正解表

細問	解答	細問	解答
ア・イ	4・5 (順不同) 完全解答	サ. シス	2. 6 1
ウ	2	セ・ソ	3・5 (順不同) 完全解答
エ	3	タ	4
オカ. キ	1 2. 8	チ	1
クケ. コ	2 5. 9		