

# 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

兵庫県立姫路西高等学校 林 宏樹

連絡先 (Email civic\_15z@yahoo.co.jp)

## 1. はじめに

2020 年、データサイエンス教育を学校全体に実装するカリキュラム開発校として、文部科学省のスーパーサイエンスハイスクールに兵庫県立姫路西高等学校（以後、姫路西）が採択された。データサイエンスを基盤とした課題研究を実施するために、データサイエンス教育を実践する。本校のデータサイエンス教育のカリキュラム開発は、兵庫県立大学と帝国データバンクとの共同開発である。

## 2. データサイエンス教育に関するカリキュラム

姫路西における 3 年間のデータサイエンス教育（以下 DS 教育）は、1 年次を「探究準備期間」、2 年次を「探究実践期間」、3 年次を「探究展開期間」と位置づけ、下記の単位数を設定している。

|       | 準備(1 年) | 実践(2 年) | 展開(3 年) |
|-------|---------|---------|---------|
| 国際理学科 | 4 単位    | 3 単位    | 2 単位    |
| 普通科   | 2 単位    | 2 単位    | 1 単位    |

国際理学科 1 年は、学校設定科目「データサイエンス研究（略称 DS 研究）（4 単位）」を実施し、それによって「社会と情報（1 単位）」「課題研究（1 単位）」「総合的な探究の時間（2 単位）」の代替とする。

普通科 1 年は、学校設定科目「データサイエンス探究（略称 DS 探究）（2 単位）」を実施し、それによって「社会と情報（1 単位）」「総合的な探究の時間（1 単位）」の代替とする。

## 3. データサイエンス教育の取組内容

DS 教育を実践していく上で、文部科学省（2018）『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、本校で実施する DS 教育では、以下のような 5 つのプロセスに細分化・具体化した。

図 1 のように、「問題－計画」を合わせ「課題の定式化」、「データ」を「データ収集」・「データ研磨」に分割、「分析」を「データ解析」、「結論」を「結果の考察」と定め、5 つのプロセスとした。

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」「情報Ⅱ」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含めている。



図 1 姫路西高校で定めた DS サイクル

## 4. 年間計画

下表の通り、前半は DS サイクルの「課題の定式化」「データ収集」「データ研磨」を重点的に学び、後半は「データ解析」「結果の考察」を重点的に学び、プロセスに応じた授業を実践し、成果として、発表会を実施した。

|           |                                 |
|-----------|---------------------------------|
| 4 月～5 月   | 一連の研究プロセスの理解<br>データサイエンス入門講座の実施 |
| 6 月～7 月   | データ分析による課題発見力の育成                |
| 7 月       | データ分析Ⅰ発表会                       |
| 8 月～12 月  | データ解析による課題発見力の育成                |
| 11 月・12 月 | データ解析Ⅰ発表会                       |
| 1 月～3 月   | DS サイクルの振り返り<br>新価値創造講座         |

DS 探究（普通科）は兵庫県統計グラフコンクール<sup>[1]</sup>に 236 名が参加し、DS 研究（国際理学科）は統計データ分析コンペティション<sup>[2]</sup>・第 21 回日経ストックリーグに参加し、87%の生徒が外部コンテストに参加した。

## 5. DS サイクルの授業内容

### (1) 課題の定式化

図 2 のように、研究テーマから課題発見に向けて、抽象的な内容から具体的な内容への思考回路を学ぶ授業で展開し、グループワーク等から課題発見に導いた。

【DS サイクル】 課題の定式化 Problem elicitation And formulation

- ① 研究テーマを決めよう。【抽象的】
- ② ①に対して、どんな理想をもっているか考えよう。
- ③ ①の「現実」について考えよう。（調査しよう。）【具体的】
- ④ ②と③のギャップから問題が見つけよう。【抽象的】
- ⑤ ④の問題から自分が解決したい課題を見つけよう。【具体的】

・言葉の意味や定義は定まっているか？  
→ 「ニート」って何？  
きちんと調べる（辞書や国・地方自治体のホームページをみる）

・意味や定義が不明確な場合は、自分自身で仮に定義する  
→ 研究発表の際の表現を確認しよう。

・似ている言葉の違いを比較する。  
→ 幼稚園と保育園の違いは何でしょう

図 2 課題の定式化の授業スライド例

## (2) データ収集

オープンデータの取得方法やダウンロードの方法等基本的な内容や、図3のように「高等学校データサイエンス教育研究会」にあるオープンデータの集約したページの紹介等を実施した。また、e-Stat(政府統計の総合窓口)<sup>[3]</sup>・RESAS(地域経済分析システム)<sup>[4]</sup>・SSDSE(教育用標準データセット)<sup>[5]</sup>などの紹介、使い方を授業で行い、次の「データ研磨」のプロセスにつなげた。



図3 データ収集の授業スライド例

## (3) データ研磨 (データ整理)

データ収集したビッグデータを整理・整形するために、下記のようなデータ研磨を行うための技能を身につけさせた。

- ・エクセルの基本操作
- ・度数分布表、ピボットテーブルの作成
- ・データの視覚化(度数分布表の作成、ヒストグラム、散布図、箱ひげ図の描画)
- ・代表値や標準偏差・相関係数等の求値方法
- ・絶対参照、相対参照等のエクセル操作
- ・外れ値、異常値の読み取り
- ・推定値と実値の差分によるデータ加工 等

なお、令和2年度の実施により、本校の実態としては想定外にエクセルの操作の指導に時間がかかることを明記しておく。

## (4) データ解析

課題解決のためのデータ解析の基本的な内容から手法について、下記のように授業を実施した。

- ・グラフの読み取り
- ・標準偏差、相関係数の理解
- ・検定入門 (p 値の活用方法)
- ・回帰分析法入門 (回帰直線・決定係数の理解等)
- ・重回帰分析法を利用したモデル構築入門
- ・クラスタリング入門 (k-means 法)
- ・連携企業によるデータ分析教材でのデータ解析 (時系列分析データの読み取り、季節要因分析の見方・考え方等)

## (5) 結果の考察

先行研究や各種生徒が参加する過去の受賞例の分析等により、データの扱い方、データとデータの組み合わせ方等をグループ討議により検証する授業を実施した。

## 6. 令和2年度におけるデータサイエンスを基盤とした探究活動の成果

令和2年度、「DS 探究」では普通科は兵庫県統計グラフコンクールに236名参加し、入選3作品・佳作5作品の結果である。また、「DS 研究」では統計データ分析コンペティションに4作品を提出し、優秀賞と特別賞を受賞した。

## 7. 大学入学共通テスト「情報」施策問題による「DS 研究」の成果の検証

大学入学共通テスト「情報」施策問題を参考に、本年度の取組に照らし合わせて加工したテストを「DS 研究」の受講生徒(39名)に実施した。

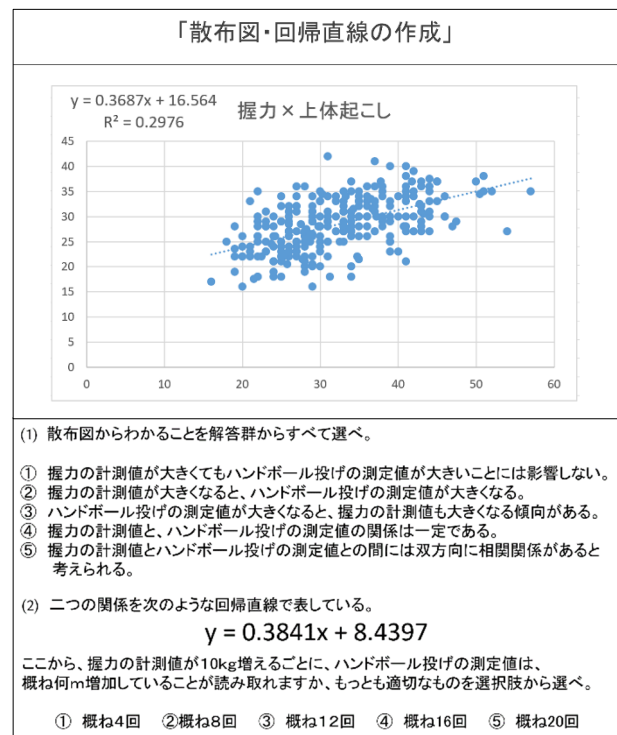


図4 確認テスト

確認テスト(図4)を実施した。

(1)における生徒(39名)の回答は、①6名、②10名、③35名、④2名、⑤33名(複数選択可)となった。⑤の回答が多かったことから、やはりまだ統計的なデータの見方が浸透されていないことが読み取れる。研究における数値からの結論の導き方も断定的な言い回しが多く、可視化、数値の読み取り等を順序立ててデータ解析していく指導を心がけていく。

(2)における生徒の回答は、①30名、②4名、③3名、④1名、⑤1名という結果であった。「回帰分析法入門」の講義において、回帰直線の表示、読み取り等の理解を目指す授業は実施したが、回帰式の理解・活用方法を深めるような内容に踏み込めていないことがわかった。次年度の「探究実践期間」においては、重回帰分析法の予測モデルの構築に関する授業を通して回帰分析法の理解の充実を目指していく。

## 8. 「DS 探究」「DS 研究」の評価基準について

本授業における評価は以下のような項目により数値化した。

### 【知識及び技能】

- ・表計算ソフトウェアを利用したデータ解析テストの得点結果
- ・タイピングテストの得点結果

### 【思考力・判断力・表現力】

- ・情報モラルに関するレポート点
- ・表計算ソフトウェアを利用したデータの可視化テストの得点結果
- ・ルーブリックによる研究発表に関する得点

### 【学びに向かう力・人間性】

- ・提出に関する評価
- ・研究における討議力や研究の態度に関する評価
- ・DS サイクルを探究活動に活用する態度に関する評価

以下に、「ルーブリックによる研究発表に関する得点[8]」について詳しく記す。

本校は、データサイエンス教育に特化したルーブリックを作成し、評価に活用した。なお、ルーブリック評価に関しては、スーパーサイエンスハイスクールの「知」の検証プロジェクトとして、奥村好美准教授（兵庫教育大学）・宮田佳緒里准教授（兵庫教育大学）と定性的評価と定量的評価の観点から共同研究している。

2020 年 11 月 19 日、DS 研究における「データ解析発表会」を実施した。1つの研究に対して、4名の教員が図6のデータサイエンス教員評価表を基に5段階評価をつけた。

|      |                               |
|------|-------------------------------|
| 評価点1 | データをグラフで可視化できていない             |
| 評価点2 | データをグラフで可視化できている              |
| 評価点3 | ・言葉の定義が定まっている<br>・課題が細分化されている |
| 評価点4 | 必要なデータが収集できている                |
| 評価点5 | グラフの特徴的な部分に関する説明がされている        |

図6 データサイエンス教員評価表

結果、評価点1は0%、評価点2は40%、評価点3は23%、評価点4は30%、評価点5は8%という割合になった。また、同発表会では「課題発見力」「課題解決力」「発信力」の3つの観点のルーブリックによる5段階評価も同時に実施した。この3つの観点と図6の評価点のそれぞれの相関を調べると、「課題解決力」との相関係数が0.94となり、強い正の相関が見受けられた。

### 【考察】

この教員評価表の評価と、「課題解決力」に強い正の相関があることは、DS サイクルにおける「課題の定式化」が、生徒の研究指導において最もハードルが

高いことを示している。評価点3にある言葉の定義や課題の細分化は、社会情勢の認識不足や語彙不足等、現実的な実態を深く考察したことがない生徒が多いという実態を表していると考えられる。そのため、抽象的な課題設定をした生徒は、データ解析のためのデータ収集で行き詰まることが多い。逆に、課題が細分化された研究は、内容に適したデータ収集がしやすいため、データ解析結果も仮説検証しやすくなる傾向が見受けられた。

## 9. 生徒アンケートによる「DS 探究」「DS 研究」の振り返り

本取組におけるアンケート調査を実施した。自分の能力についての質問として、DS サイクルのプロセスごとに身につく力を、「5. 十分ある」～「1. 不足している」の5件法で調査した結果が図5である。

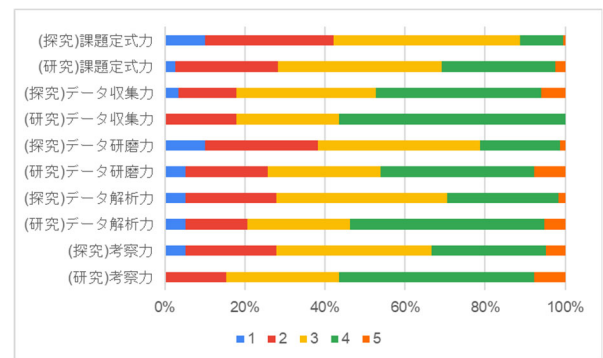


図5 DS サイクルの生徒による自己評価

### (1) 「探究(普通科)」と「研究(国際理学科)」の比較

高評価(4・5 評価点)の割合をみると、5つの力のすべてにおいて、国際理学科の方が高い。単位数の違いにより、普通科に比べ、連携企業によるデータ分析教材の授業や、研究指導においても綿密な指導ができたため、想定通りの結果である。

### (2) DS サイクルのプロセスの比較

普通科・国際理学科ともに、課題定式力が最も低いことがわかる。生徒の意識においても、「課題の定式化」が難しいと感じており、来年度は課題の定式化に関する授業の工夫をしていく。

## 10. 今年度の振り返りと今後の展望

(1) 本年度の「探究準備期」では、DS サイクルの「データ収集」「データ研磨」を重点的に実施した。想定以上に「データ研磨」のエクセル操作の指導に時間がかかった。また、定着の悪い内容とそうでない内容の区別が計画にできなかった。今後、「データ収集」「データ研磨」の指導内容を精査し、効率的に知識・技能を習得する授業展開に改良していく。また、授業計画では指導の順序を工夫し、「課題の定式化」により多くの時間をかけられるように計画を立てていく。

(2) 授業時間的には仕方がないが、「データ解析」の手法においては、入門的な内容しかできず、実際、生徒の探究活動で統計手法を扱った内容はほとんど見受けられなかった。しかしながら、本年度の取組において、これでよいと考えている。むしろ、生徒の探究の内容をみると、代表値・標準偏差・相関係数・度数分布表・ヒストグラム等の数学Ⅰにおける内容を丁寧に扱い、分析する土台作りを「探究準備期間」にする必要がある。今後は、統計手法は「探究実践期間」に重点的に扱い、「探究準備期間」はデータを可視化・整理し、データを眺める習慣を身につけさせる。

(3) データサイエンスに特化した評価に関しては、本年度の取組から「知」の検証プロジェクト（8に記載）により、下記のルーブリックを作成した。なお、字体の色はDSサイクルとの関連を示し、各プロセスの評価1から段階的に変化した場合だけ色を変えている。（黄緑色はデータ収集・紫色はデータ研磨・水色はデータ解析・オレンジ色は結果の考察・赤色は創造力）

〔「課題発見力」に関する教員評価の記述語〕

| 評価 1                                                                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>必要なデータがない</li> <li>主観的な読み取りである</li> <li>仮説間の関連が全くない</li> </ul>                                                                                                       |
| 評価 2                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>必要なデータはあるが、適切に加工されていない</li> <li>主観的な読み取りである</li> <li>データと課題・仮説が対応していない</li> <li>仮説間の関連が全くない</li> </ul>                                                               |
| 評価 3                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>必要なデータの収集・抽出・加工がされている</li> <li>客観的な読み取りである</li> <li>データと課題・仮説につながりがある</li> <li>仮説間につながりがある</li> </ul>                                                                |
| 評価 4                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>必要なデータ（自分たちで適切なデータの採取を含む）の収集・抽出・加工</li> <li>客観的な読み取りである</li> <li>データと課題・仮説間に適切なつながりがある</li> <li>複数のグラフや表から独自性のある課題発見・仮説設定である</li> </ul>                              |
| 評価 5                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>既存のデータ収集に加え、自分たちで適切なデータを採取するなどの工夫と加工</li> <li>客観的な読み取りと適切に統計手法を活用した分析が実施されている</li> <li>データと課題・仮説間に確かに適切なつながりがある</li> <li>複数のグラフや表から独自性・新規性のある課題発見・仮説設定である</li> </ul> |

令和3年度はこのルーブリックで評価し、データサイエンスに基づく「課題発見力」を生徒の実態、活動との整合性と照らし合わせて検証していく。

(4) 来年度「探究実践期間」では、「準備期間」の取組を振り返り、1つ1つのプロセスの定着とともに、「データ解析」プロセスにおける統計分析手法を習得させ、データ解析力を重点的に養っていく。「情報Ⅱ」に関する内容を重点的に指導し、機械学習の仕組みを学び、AIを活用するための知識を身につけ、活用する技能を習得させていく。そして、「結果の考察」のプロセスから新たな価値を見出す創造力の育成を目指し、AI時代を生き抜く生徒の育成を行う。

## 参考文献

- [1] 兵庫県統計グラフコンクール  
[https://web.pref.hyogo.lg.jp/kk11/ac08\\_1\\_000000189.html](https://web.pref.hyogo.lg.jp/kk11/ac08_1_000000189.html)
  - [2] 統計データ分析コンペティション  
<https://www.nstac.go.jp/statcompe/>
  - [3] e-stat 政府統計の総合窓口 <https://www.e-stat.go.jp/>
  - [4] RESAS 地域経済分析システム  
<https://resas.go.jp/#/13/13101>
  - [5] 統計データ分析コンペティション  
<https://www.nstac.go.jp/statcompe/>
  - [6] 兵庫県統計グラフコンクール入賞作品掲載  
[https://web.pref.hyogo.lg.jp/kk11/oshirase-sougoude-ta/toukeigrafkonkuru26\\_pc\\_nyuusen.html](https://web.pref.hyogo.lg.jp/kk11/oshirase-sougoude-ta/toukeigrafkonkuru26_pc_nyuusen.html)
  - [7] 統計データ分析コンペティション受賞論文掲載  
<https://www.nstac.go.jp/statcompe/award.html>
  - [8] 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、評価基準の作成」IPSJ\_SSS2020022.pdf、2020-12-12
- ・長尾篤志（2019.9.4）『教育課程部会 長尾主任視学官発表資料「理数探究」の充実とSTEAM教育について』
- ・文部科学省（2018）『高等学校学習指導要領 解説 理数編』