

令和3年度SSH交流支援事業

データサイエンティストへの道

～全国高等学校データサイエンス教員研修会～



主催：兵庫県立姫路西高等学校

1. 目的

近年、データ活用人材を育成する必要性が増し、高等学校の令和4年度から実施される学習指導要領においては、データサイエンス（以下、DS）に関わる学習内容が多く含まれている。また、探究活動においてDSの活用によって探究内容が深化されていく点は、本校の実践でも実証されている。そのような現状を踏まえ、DS教育に関する高等学校教員の指導力向上を目的とする「全国高等学校DS教員研修会」を開催する。

2. 目標

目的を達成するため、次の3つの目標を掲げる。

- (1) DSに関する知見を広げることにより、指導力を向上させる。
- (2) DSへの抵抗感をなくすことで、DSについて指導する意欲を高める。
- (3) 産学連携・高大接続により、実社会や大学とのつながりを意識した高等学校での教科指導の改善をはかる。

3. 主催

兵庫県立姫路西高等学校

なお、本事業は、国立研究開発法人科学技術振興機構における令和3年度SSH交流支援事業による支援を受けている。

4. 日程・実施回数

- | | | | |
|-----|-----------|-------------|----------------------|
| 第1回 | 9月10日（金） | 15：30～16：50 | タイトル「コロナ×データサイエンス」 |
| 第2回 | 9月24日（金） | 15：30～16：50 | タイトル「医療×音声×データサイエンス」 |
| 第3回 | 10月15日（金） | 15：30～16：50 | タイトル「広告×データサイエンス」 |
| 第4回 | 10月29日（金） | 15：30～16：50 | タイトル「音楽×AI×データサイエンス」 |
| 第5回 | 11月17日（水） | 15：30～16：50 | タイトル「スポーツ×データサイエンス」 |
| 第6回 | 12月17日（金） | 15：30～16：50 | タイトル「情報教育×データサイエンス」 |
- 全6回の実施をまとめ、本校教職員、他校参加者へ還元している。

4. 実施方法

オンラインによる講義・講演

5. 研修会参加者

本校職員を含めた全国高等学校教員

6. 研修内容

第1回 9月10日(金) 15:30~16:50	
研修タイトル 「コロナ×データサイエンス」	
講師 鈴木 和幸 (国立大学法人電気通信大学名誉教授・特任教授)	
日本品質管理学会会長, 日本信頼性学会会長, 積水化学工業(株)社外役員, 統計数理研究所客員教授。 品質管理とは顧客と社会の満足を目的に, データに基づき意思決定と行動を行うことであるという概念に基づき, 品質管理に50年間携ってきた。	
研修内容 データサイエンスもAIも目的ではなく手段・方法である。大事なことは, これらを用いて如何に問題を解決し, 社会に貢献できるか, そのプロセス(文法)を学び, これを実践に役立たせることである。データサイエンスとは何か。なぜデータサイエンスが必要か。データサイエンスの前提となる科学技術の背景とともに, コロナ感染阻止を目的に, データに基づく意思決定の重要性を知ってもらう。	

第2回 9月24日(金) 15:30~16:50	
研修タイトル 「医療×音声×データサイエンス」	
講師 菅 由紀子 (株式会社 Rejoui 代表取締役)	
データサイエンティスト協会スキル定義委員, 関西学院大学大学院経営戦略研究科兼任講師。HBMS 県立広島大学大学院非常勤講師。 2016年9月(株)Rejouiを設立し, データサイエンティスト育成事業・独自の機械学習アルゴリズムを活用した学習サービス事業を展開。	
研修内容 医療分野におけるデータの利活用はたいへんめざましく, 世界中で様々な活用が行われております。発表では当社の請け負った診察室における音声データを用いたデータ解析(音声認識, 言語処理を用いた統計解析と機械学習の実行)事例をご紹介します。	

第3回 10月15日(金) 15:30~16:50	
研修タイトル 「広告×データサイエンス」	
講師 原野 朱加 (株式会社野村総合研究所主任コンサルタント) データサイエンティスト協会スキル定義委員。 野村総合研究所ではマーケティングコンサルタントとして従事し、ブランド戦略立案や広告効果測定、市場調査、事業予測など、データを活用したマーケティング活動支援を専門とする。	
研修内容 データサイエンスは日常的に目にする広告に様々なアプローチで貢献している。デジタル広告などのわかりやすい事例だけでなく、テレビCMなどのマス広告におけるデータ活用事例を紹介する。また、広告の話の起点に、マーケティング全般にも少し目を広げた事例発表を行う。	

第4回 10月29日(金) 15:30~16:50	
研修タイトル 「音楽×AI×データサイエンス」	
講師 深山 覚 (国立研究開発法人産業技術総合研究所主任研究員) 情報理工学博士。日本学術振興会特別研究員(DC2)、 産業技術総合研究所研究員を経て、現在同研究所主任研究員。 専門はメディア情報学・音楽情報処理。	
研修内容 私たちが日々楽しむ身近な音楽をデータサイエンス・人工知能の視点で捉えると、どのようなことができるかを紹介する。音楽のデータとは何か、音楽のデータ分析をすると何ができるのかを概説した上で、最新の音楽情報処理の研究事例を自身の研究をまじえて紹介する。また新しく音楽の分析を始めたい人に向けて、研究目的で使用可能なデータセット・ツールの紹介を行う。	

第5回 11月17日(水) 15:30~16:50	
研修タイトル 「スポーツ×データサイエンス」	
<p>講師</p> <p>廣澤 聖士 (日本スポーツ振興センタースポーツ科学部)</p> <p>独立行政法人日本スポーツ振興センターハイパフォーマンススポーツセンター・国立スポーツ科学センター。専門分野・映像分析。フィギュアスケートバジジテストシングル3級。</p> <p>ハイパフォーマンスサポート事業ターゲット種目に対する映像・分析サポート (ショートトラックスピードスケート) を行う。</p>	
<p>研修内容</p> <p>今日ではスポーツ界でも様々な文脈でデータ活用が行われるようになっている。スポーツのデータ活用に関して幅広い分野からのオーバービューを取り扱い、業界のデータ活用のトレンドを紹介する。後半は、講演者の専門競技であるアイススケートを題材に、パフォーマンス (競技力) 向上の文脈でのデータ活用を取り扱い、現場の強化の事例や基本的なデータサイエンスの手法を適用した研究事例を発表する。</p>	

第6回 12月17日(金) 15:30~16:50	
研修タイトル 「情報教育×データサイエンス」	
<p>講師</p> <p>鹿野 利春 (京都精華大学メディア表現学部教授)</p> <p>国立教育政策研究所教育課程研究センターにて教育課程調査官、文部科学省初等中等教育局情報教育・外国語教育課情報教育振興室教科調査官として勤務。情報科の新学習指導要領、情報教育、GIGA スクール構想などを担当。現在は、文部科学省初等中等教育局視学委員。</p>	
<p>研修内容</p> <p>小中高で行われている統計教育、高校情報科で行う統計・データサイエンス教育、高校に期待するカリキュラム・マネジメント、1人1台情報端末とデータサイエンス教育の関係等、高等学校教育全般においてデータサイエンスの役割や重要性を理解し、データサイエンス教育に携わる有用性を解説する。</p>	

1. 兵庫県立姫路西高等学校における DS 教育と研修会との関連

本研修会では、兵庫県立姫路西高等学校（以下、姫路西）における DS 教育の根幹となる統計的探究プロセス、PPDAC サイクルという研究フローに基づいて、様々な分野のデータサイエンスの専門家に実践事例や分析手法を講演していただいた（図 1, 2）。

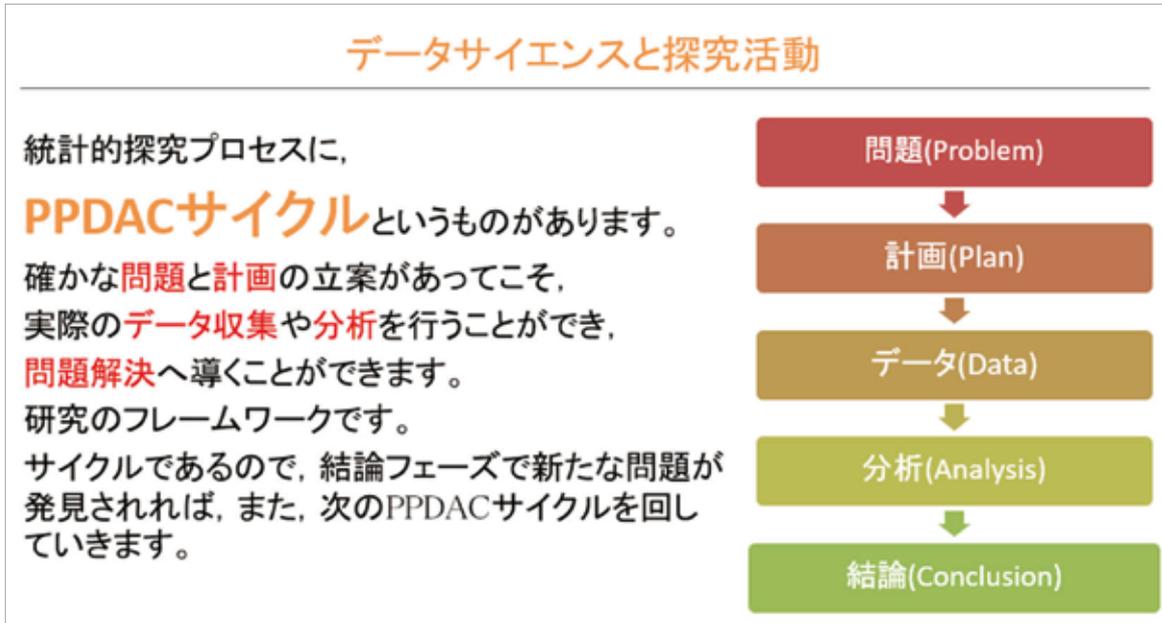


図 1 PPDAC サイクルについて

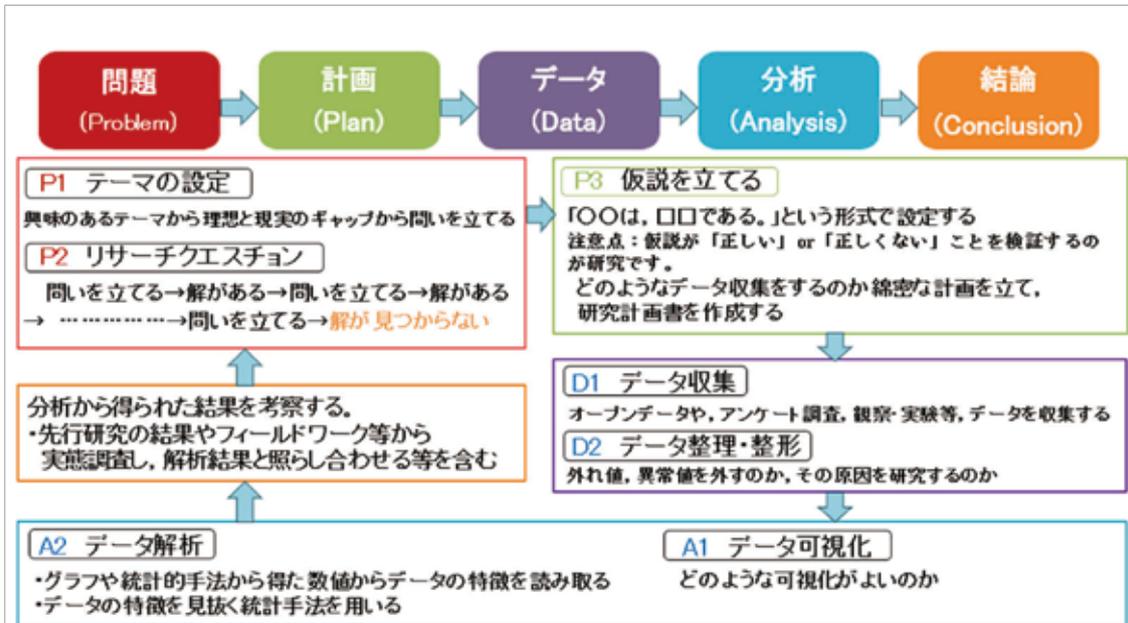


図 2 PPDAC を細分化した姫路西の指導事例

2. 研修会の内容（抜粋）

(1) 第1回研修会 「コロナ×データサイエンス」

アメリカにおけるコロナウイルス対策に関するデータサイエンスの事例研究であった。QC ストーリー的問題解決プロセスという研究フローがある。図 4 の左側に PPDAC サイクルに該当する箇所が記載されているため、教員が指導する際には、PPDAC サイクルで指導を進めつつ、QC ストーリーに沿った助言をすることで生徒に指導しやすくなる。

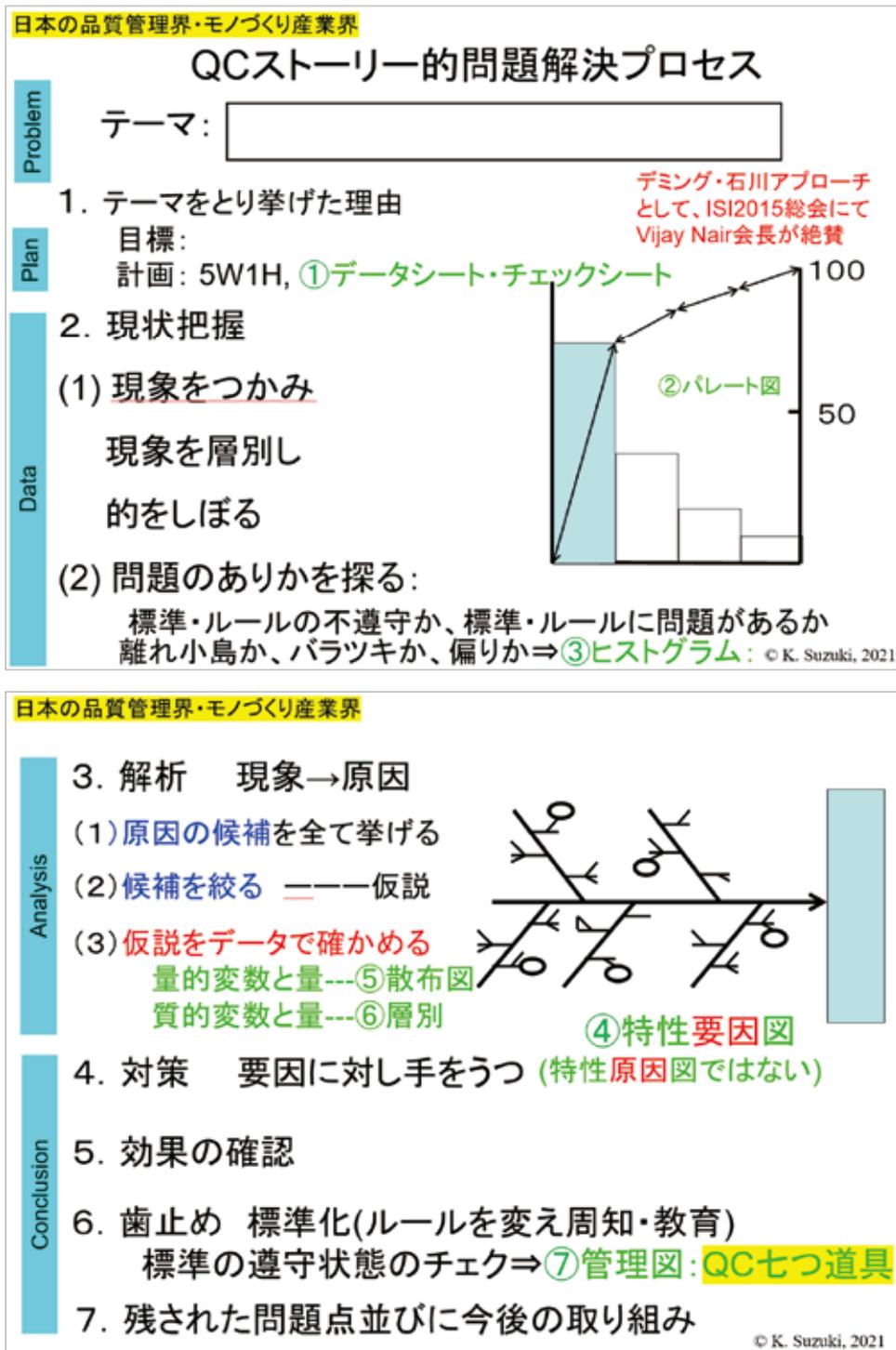


図 4 QC ストーリー的問題解決プロセスの概要

研究では、綿密な研究デザイン（計画）を立てる必要がある（図5）。どのようにデータを分類，比較するのかを見定め，データの整理・整形，可視化，解析を進めていく（図6）。データ解析の際には，グラフを比較することが有効な手段である。

【デザイン】
2020年9月30日までの全米50州とDC特別区マスク義務化情報（一部義務化の州については対人口割合）

- A 6/5前の州全体の義務化の州
- B 6/5以降の州全体の義務化の州
- △ C 州一部実施人口割合30%以上の州
- × D 州一部実施人口割合30%未満の州

の4群を考える。

【米国を取り上げる理由】
 これまでマスク着用の慣習がなく、その義務化に**当たり州毎に施策内容とそのタイミングが異なり社会実験と見做すことが可能であるため**

© K. Suzuki, Sep. 2021 47

図5 研究デザインの事例

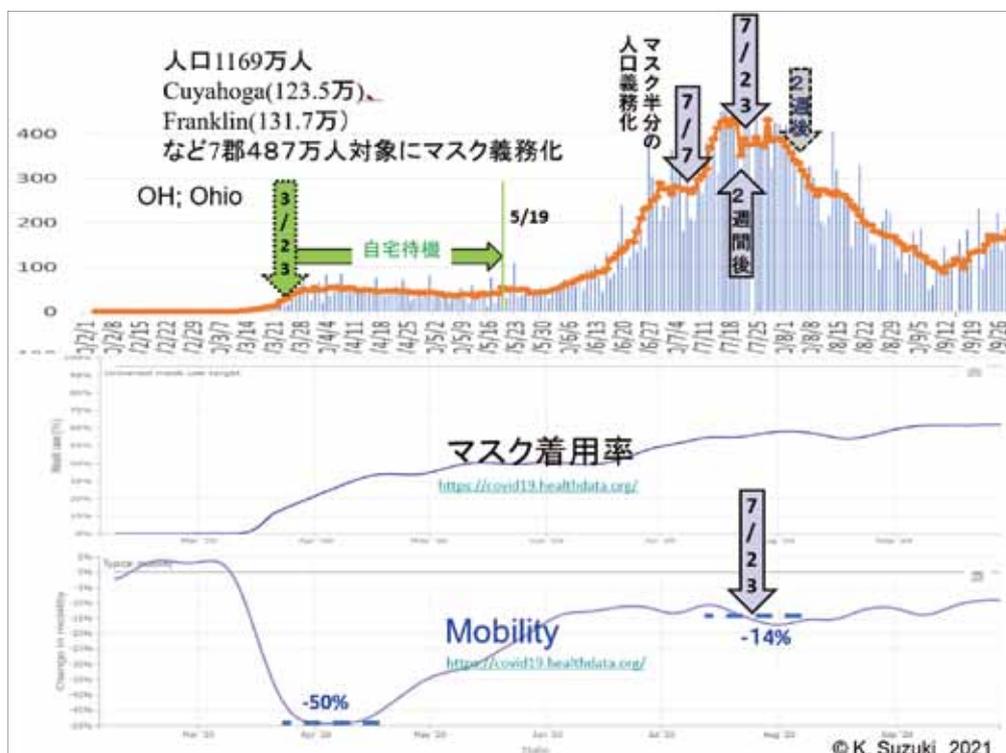


図6 データ解析がしやすくなるデータの整理・整形，可視化の事例

事例発表では、図8のようにデータを示されたおかげで、非常に伝わりやすい可視化であったが、生徒が研究をすると図7のように「とりあえず散布図をかく」ことで可視化を行い、データ解析できないという事例がある。図8のように、色を分けて可視化することで、データが表す本質を見抜くことができる可視化となる事例であった。

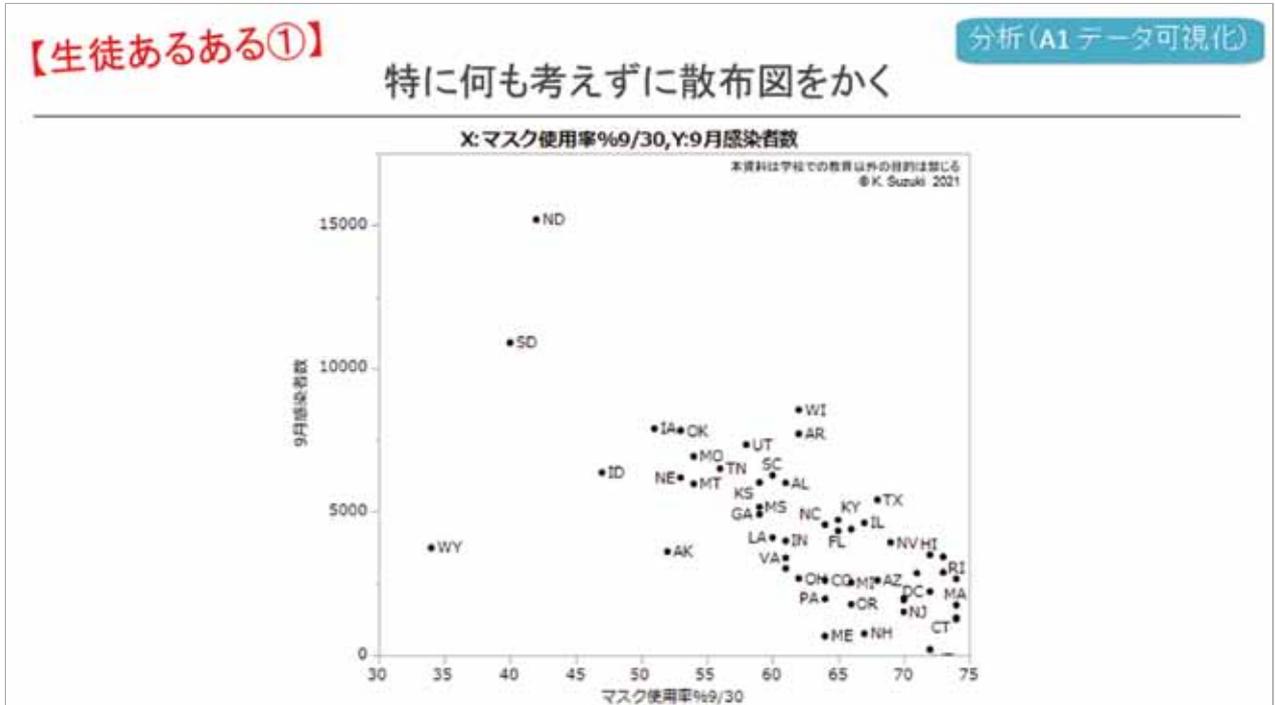


図7 工夫せずにかいた散布図

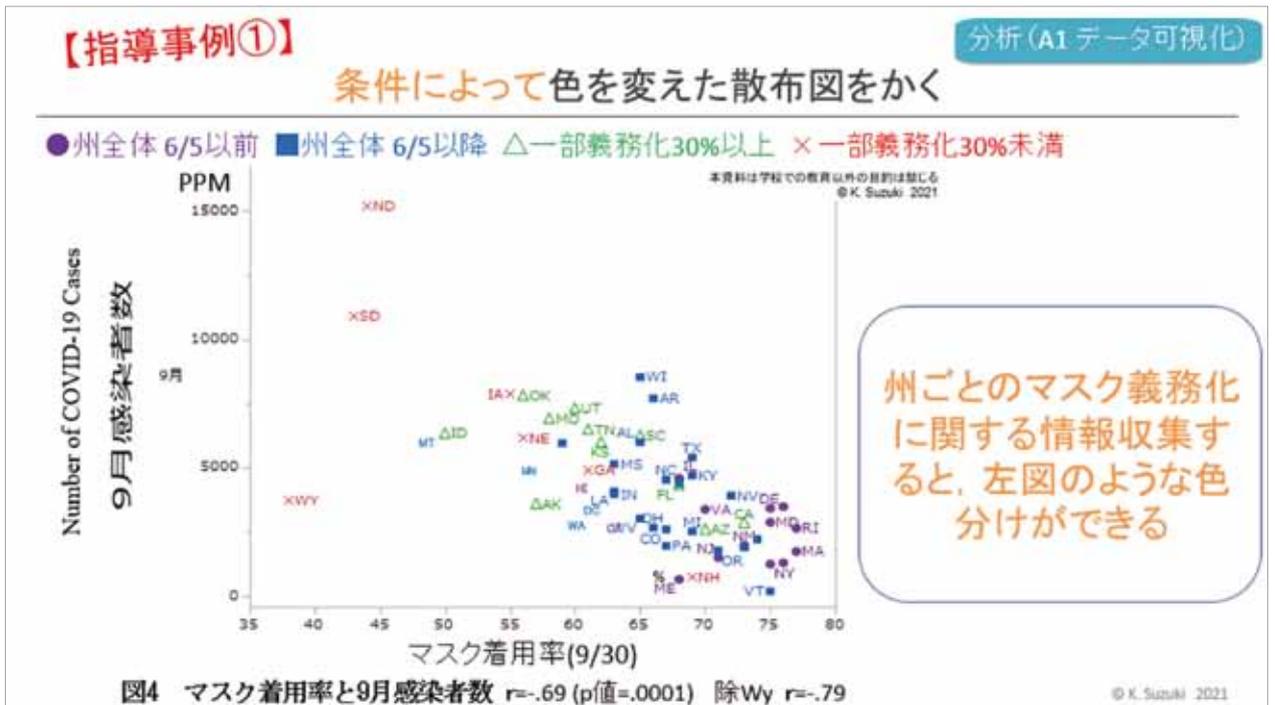


図8 図7を色分けした散布図

生徒の研究では、図9のように散布図を並べて比較することがあるが、このままではデータ解析できないことに気づかせたい。図9では、2つの散布図は縦軸のスケールが異なっており、データを読み取る際には、図10のように軸のスケールを揃えるように指導することが重要である。

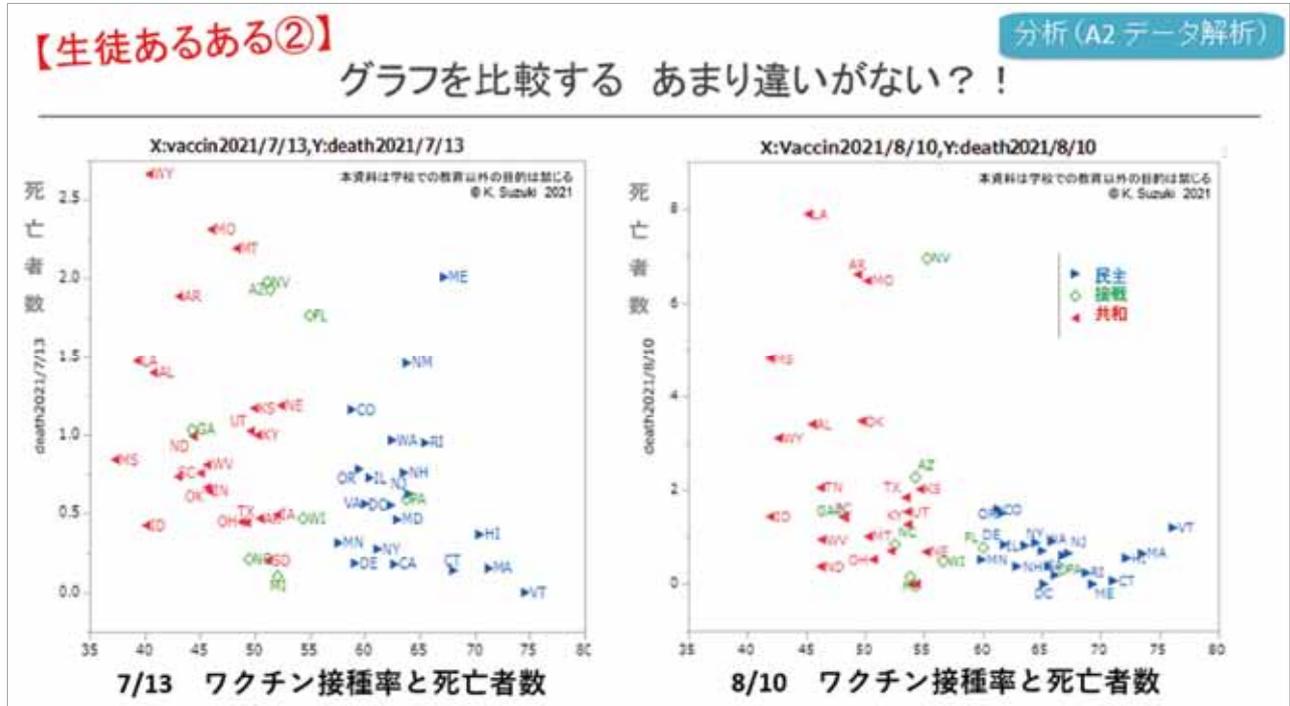


図9 異なる時期のデータを並べた散布図

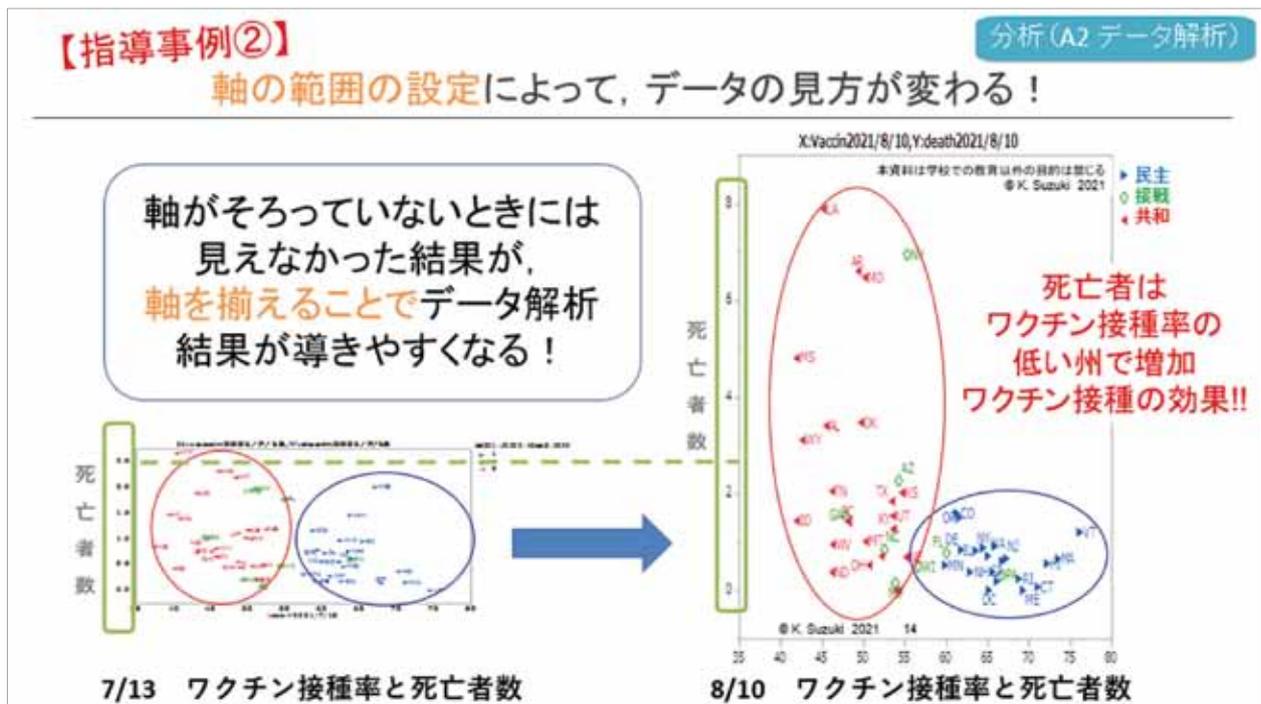


図10 軸を揃えて並べた散布図

(2) 第2回研修会 「医療×音声×データサイエンス」

医療における音声データを分析するデータサイエンス事例発表であった。

専門家が実施するPPDACサイクルの一事例が図11である。綿密な研究計画の設計が重要であることがわかる。データのプロセスでは、データ収集だけでなく、整理・整形となる前処理を丁寧に行うことで、次のプロセスの分析・考察につながる(図12)。

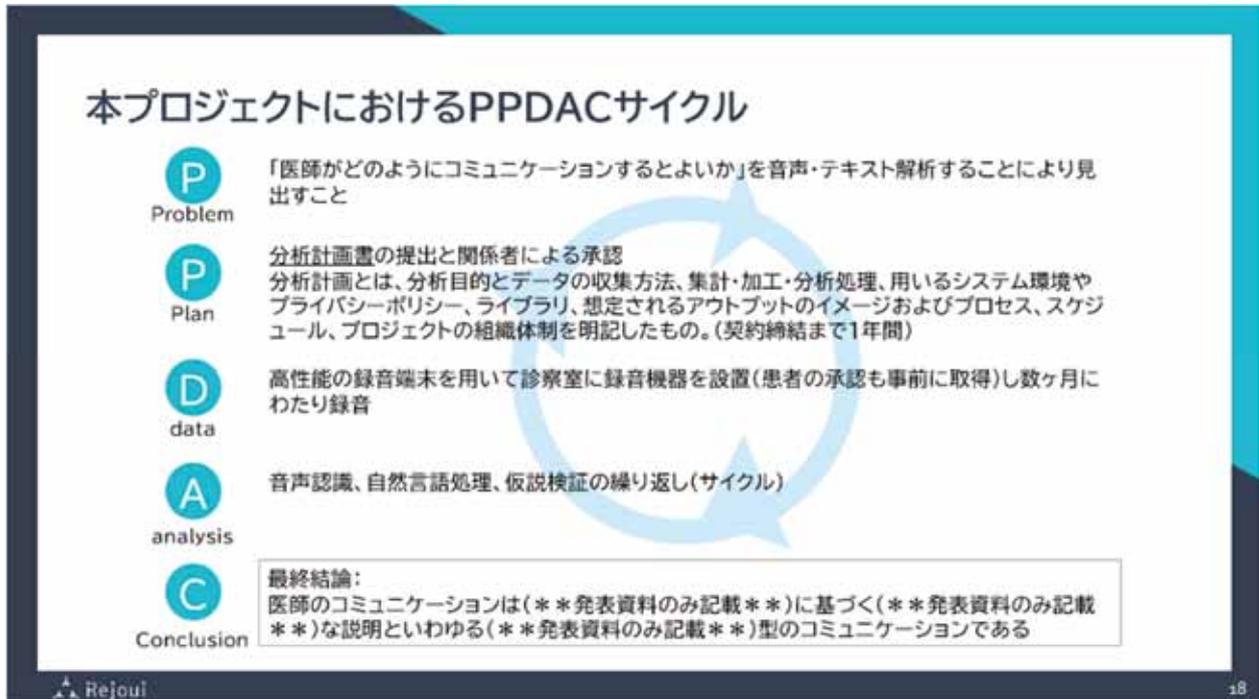


図11 専門家が行うデータサイエンスに関するPPDACサイクルの事例



図12 専門家が実施するデータの前処理のフロー

図 13 のように音声データを処理する手段としても様々あり、データを解析した手段に応じた形に変換することが重要である。また、研究は PPDAC サイクルを 1 周まわして完了するのではなく、1 周目の結論から新たな問題を見出し、サイクルを何回もまわすことで、研究が深まる (図 14)。

音声認識とは

音声認識とは、音声や音楽などをデジタルデータに変換し、機械が扱いやすいデータにした後に様々な機械学習のモデルを用いて音声データから文章・単語を認識させる技術の総称。

プロセス: デジタル化 → 特徴量抽出 → 音声認識

活用例:

- 本人認証: 音声認識技術を用いて本人を確認する。
- テキスト解析: 音声認識されたテキストを解析する。例: 「メンバーは、医療業界経験者と広告コミュニケーション業界経験者で構成されています。」「元気な、美しく、長生きすることこそ、高齢化が進む」
- 感情分析: 音声認識されたテキストから感情を分析する。

図 13 音声データの処理する方法

医療現場の音声解析 -4_2

PPDAC サイクル: P (Problem) → P (Problem) → D (Data) → A (Analysis) → C (Conclusion)

Conclusion 音声認識を上手く行うことが出来ない

考えられる原因

- ・周囲の雑音(窓を開けている時の外音・機械音・打鍵音・受付の声など)
- ・早口
- ・不明瞭
- ・クロストーク
- ・方言
- ・診察付き添い者の会話(患者さんが子供・高齢者の場合)

Problem1 音声認識の精度が低い(上手く行うことが出来ない)

Problem2 テキスト化を間に合わせなければならない

※事例の数値は全てダミーです 27

図 14 1 周目の PPDAC の結論から次の PPDAC をまわす

図 15 のように、＜問題→計画＞を繰り返すことで、解明したい事柄に近づいていく。2 周目の PPDAC になると、1 周目に比べ、非常に具体的な処理が増えている。音声のデータに関する研究は、データ収集が難しい場面が多い。また、音声データの分析法について、フーリエ変換などの高度な手法を活用できるため、理科教員と協力し分析するとよい（図 16）。

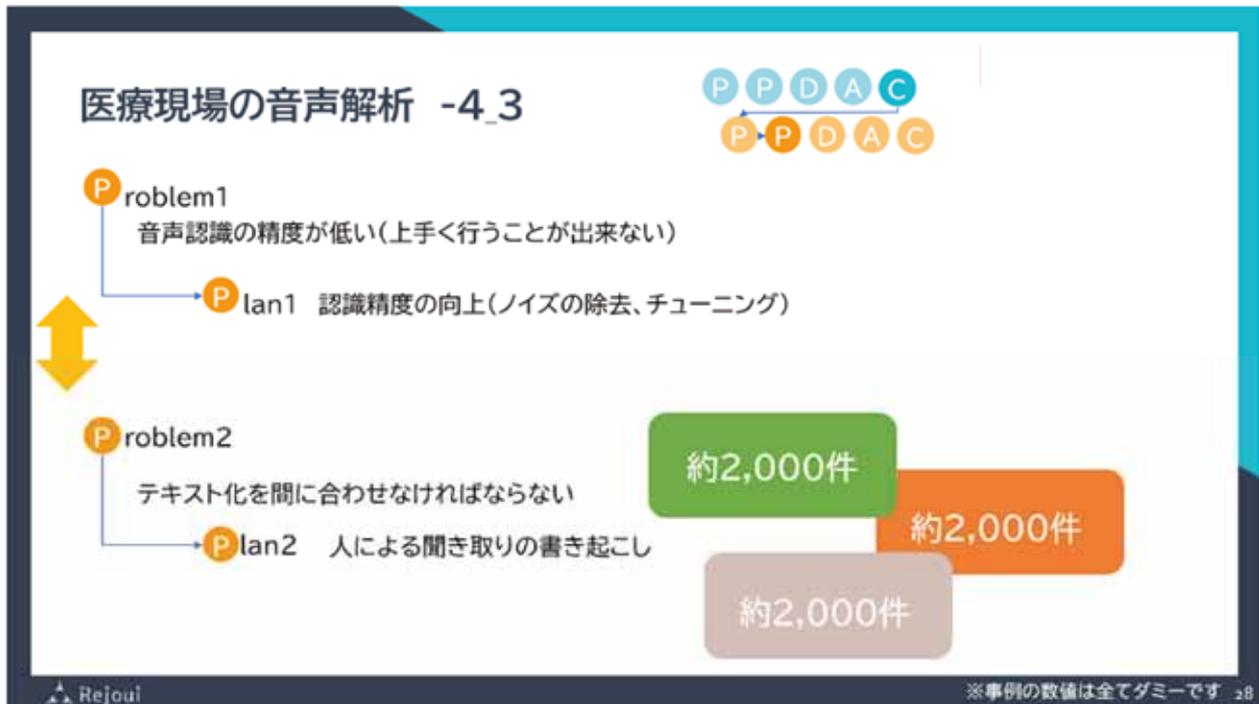


図 15 1 周目の PPDAC の結論から次の PPDAC をまわす

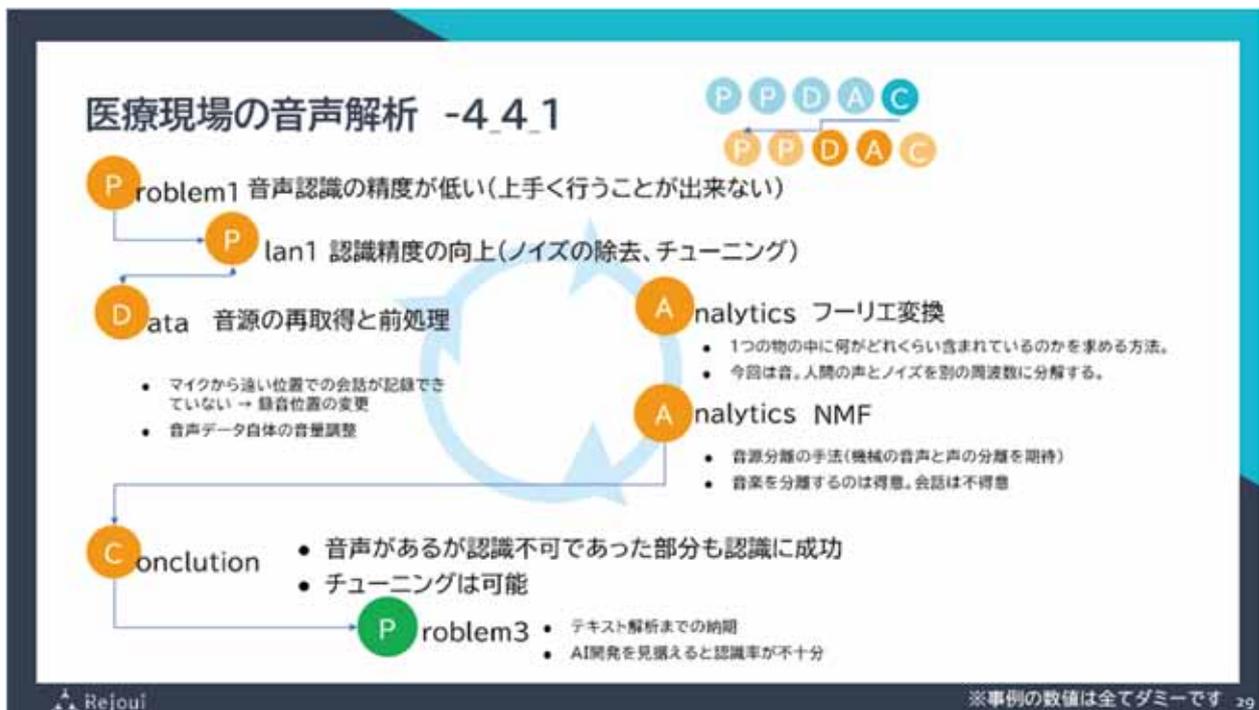


図 16 2 周目の PPDAC サイクル

2 周目の PPDAC サイクルになると高度な統計手法を用いることも多くなる。生徒の研究内容によって、図 17、図 18 のような分析手法がある。

参考:[手法説明]アソシエーション分析

アソシエーション分析は、ビジネスのマーケティングでよく使用される手法で、膨大なデータの中から統計的なパターンや、意味のある関連性を抽出するデータマイニング手法です。たくさんある商品の中から、商品間の関連性や同時性を見つけるのに有益といわれます。

信頼度(confidence)

信頼度とは、例えば、Aを買った人のうち、どれくらいの人がBも買ったかという割合のことです。
例:ショートケーキを買った人のうち、どれくらいの人がプリンも買ったか?
ショートケーキを購入した人:100人、プリンを買った人60人、両方を買った人:30人では、 $30 \div 100$ で信頼度は0.3です。

支持度(support)

支持度とは、例えば、A とBがどのくらい一緒に売れているかという割合のことです。この指標が大きいくほど全体の中でそのルールが出現する確率が高くなります。一般的に知られている支持度は一緒に売れた数を、全体の数で割るものです。
ショートケーキもプリンも両方を買った人:30人、お店に来たすべての人が200人であるとき、 $30 \div 200$ で支持度は0.15です。

リフト(lift)

リフト値とは、例えば、AとBと一緒に買った人の割合は、全体の中でBを買った人の割合よりどれだけ多いかを倍率で示したもののことです。リフト値が低い場合、AとBの関連性にあまり意味はなく、一般的にリフト値が1以上の場合は有効なルールとみなされています。

Rejou 38

図 17 アソシエーション分析の手法説明

参考:[手法説明]単語のスコアリングを行う「極性辞書」

感情極性とは、その語が一般的に良い印象を持つか(positive) 悪い印象を持つか(negative)を表した二値属性です。例えば、「良い」、「美しい」などはpositiveな極性、「悪い」、「汚い」などはnegativeな極性を持ちます。-1から+1の実数値を割り当ててあり、-1に近いほどnegative、+1に近いほどpositiveと考えられます。今回は、分析に際し東工大の高村教授が公開している極性辞書を活用しました。約5万5千語に対応した辞書です。

単語感情極性対応表(一例)

Positive スコア Top10				Negative スコア Worst10			
単語		品詞	スコア	単語		品詞	スコア
褒める	すぐれる	動詞	1	下手	へた	名詞	-0.99983
良い	よい	形容詞	0.999995	卑しい	いやしい	形容詞	-0.99986
喜ぶ	よろこぶ	動詞	0.999979	ない	ない	形容詞	-0.99988
褒める	ほめる	動詞	0.999979	濃ける	つける	動詞	-0.99995
めでたい	めでたい	形容詞	0.999645	罵る	ののしる	動詞	-0.99996
賢い	かしこい	形容詞	0.999486	ない	ない	助動詞	-1
善い	いい	形容詞	0.999314	醜い	ひどい	形容詞	-1
通す	通す	動詞	0.999295	病気	びょうき	名詞	-1
天晴	あっぱれ	名詞	0.999267	死ぬ	しぬ	動詞	-1
祝う	いわう	動詞	0.999122	悪い	わるい	形容詞	-1

形態素解析(分かち書き)から極性辞書を参照しそのまま感情スコアをスコアリングしてくれるライブラリもあります。

Rejou 40

図 18 単語のスコアリングを行う「極性辞書」

(3) 第3回研修会 「広告×データサイエンス」

PPDAC サイクルに即した広告関係におけるデータサイエンス事例発表であった。どのような分野でも、＜問題-計画＞のプロセスが最も重要である（図19）。また、シングルソースパネル調査という手法の紹介があった（図20）。

マルチメディア広告効果検証

どんな調査においても、問題～計画(仮説構築)までが非常に大切
クライアントにヒアリングし、計画・データへ落としこむ

問題 ~ **計画** プロジェクト立ち上げ時のやり取り

問題 (Problem)  クライアントへのヒアリング

- ✓ テレビCMはコストも高く、社内で不要論が上がっている
- ✓ でもなかなか今あるデータでは、デジタルの方が効率が良く見えてしまう
- ✓ テレビCMの効果や価値を定量的に示し、予算を確保したい

計画 (Plan)  コンサルタントからの提案

- ✓ テレビCMがデジタル広告よりも重要であることを定量的に示す
- ✓ そのためには広告効果を媒体ごと比較できるような調査フレームを提案（今回の場合はシングルソースパネル調査）
- ✓ 実際はこの時点で分析結果のアウトプットイメージを提示して交渉する

ポイント 実ビジネスにおいても、問題～計画策定が最も重要なプロセスです

Copyright (C) Nomura Research Institute, Ltd. All rights reserved. NRI 14

図19 問題→計画のフロー

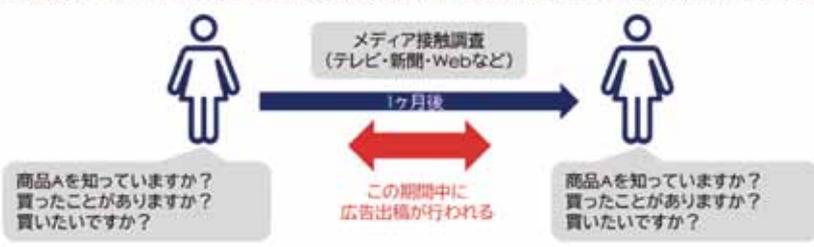
マルチメディア広告効果検証

シングルソースパネルデータを用いることで、一人のモニタの一連の
メディア接触状況と態度変容を観測することが可能

データ シングルソースパネル調査の概要

シングルソースパネル調査とは...

同一のモニタに一定期間調査協力をしてもらい、モニタの行動や意向の変化を観測する調査
広告効果測定に用いる場合は、広告出稿前と出稿後の2地点を比較し、その間のメディア接触有無によるリフトを観測する



商品Aを知っていますか?
買ったことがありますか?
買いたいですか?

商品Aを知っていますか?
買ったことがありますか?
買いたいですか?

Copyright (C) Nomura Research Institute, Ltd. All rights reserved. NRI 15

図20 シングルソースパネル調査の概要

図 21, 図 22 のような分析の流れによって, 分析手法が単純であっても「何と比較して効果があったのか」を示すことが重要であり, 結論においては, 総評を書くだけでなく, PPDAC サイクルが繰り返されていくことを踏まえた今後への示唆が重要であることがわかる。

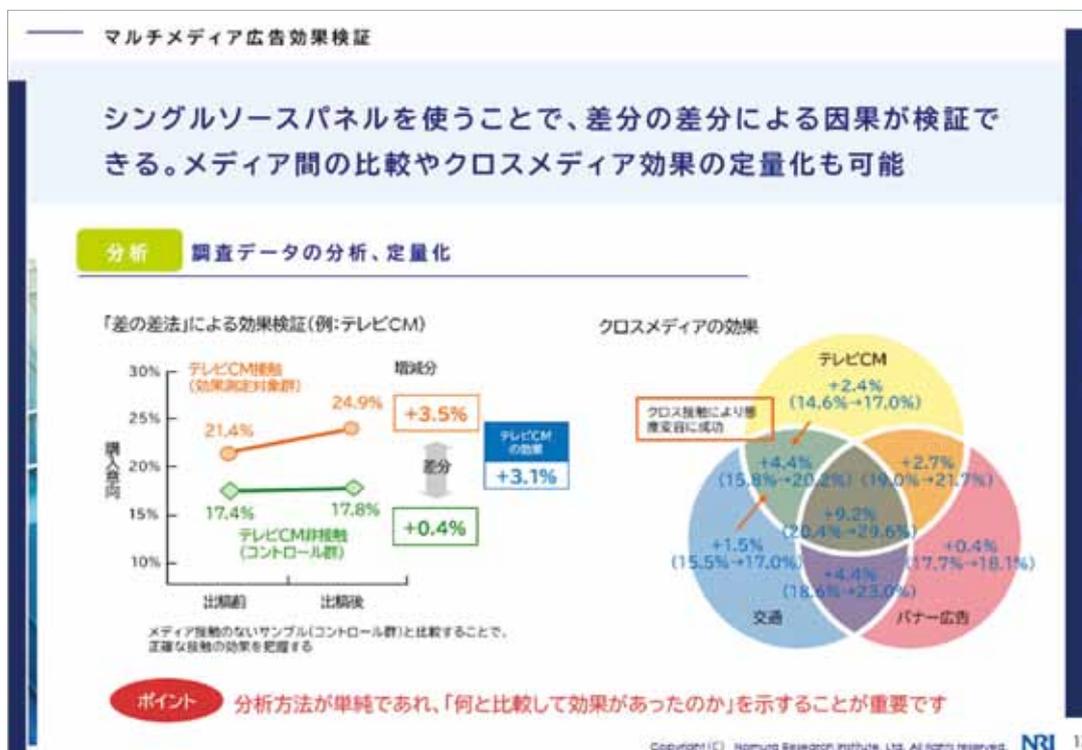


図 21 分析プロセス



図 22 結論プロセス

コレスポネンス分析, 因子分析という統計手法の紹介があった (図 23, 24)。

(補足)コレスポネンス分析とは

- ◆ コレスポネンス分析とは、クロス集計表を可視化することで調査結果の解釈を容易にする分析手法
- ◆ ブランドポジションを相対的に把握するためによく用いられる
- ◆ メリット | 特に調査項目が多いクロス集計表は特徴を把握しにくいので、可視化することで相対位置が把握しやすくなる
- ◆ デメリット | 相関行列をもとに算出されるため、回答ボリュームは反映されない点に留意が必要

コレスポネンス分析のイメージ

①行方向の分析

②列方向の分析

2つを重ね合わせる

元データ(度数分布表)

	ブランドA ブランドB	ブランドC ブランドD	安定している	信頼できる	日本を代表している	先進的である	老舗的	活気がある	文化活動を支援	社会への影響力がある	サービスが安い
ブランドA	800	250	200	100	50	800	200	150	200	200	
ブランドB	700	100	100	100	100	100	100	100	100	100	
ブランドC	600	100	100	100	100	100	100	100	100	100	
ブランドD	500	100	100	100	100	100	100	100	100	100	

※軸の抽出には主成分分析と同様のロジックが用いられる
主成分分析...次元圧縮の手法、最も分数が小さくなる2軸の合成変数を生成することができる

Copyright (C) Nomura Research Institute, Ltd. All rights reserved. NRI 25

図 23 コレスポネンス分析の概要

(補足)因子分析とは

- ◆ 因子分析とは、多数の調査項目で取得した情報を、極力情報量を落とさずにデータの構造や傾向を見出すための手法
- ◆ 潜在変数(因子)と元の調査項目との関係を解釈し、算出された因子得点を用いて特徴の可視化を行う

因子分析の概念

潜在変数(因子)

観測変数(調査項目)

要因(因子)1

要因(因子)2

要因(因子)3

因子負荷量(次ページ数表内数字)
観測変数に対して因子がどれくらいの影響を与えているかを示したもの

調査項目1 (安定している) 0.729

調査項目2 (信頼できる) 0.724

調査項目3 (日本を代表している)

調査項目4 (先進的である)

調査項目5 (老舗的)

調査項目6 (活気がある)

調査項目7 (文化活動を支援)

調査項目8 (社会への影響力がある)

調査項目9 (サービスが安い)

因子得点の意味

因子得点は各項目を因子負荷量を反映して集約したスコア
= 潜在変数を数値化したものである
したがって、この因子得点を集計することで、各企業の特性
や属性の特性の概要を把握することができる

回答データと因子得点のイメージ

モニタID	回答ブランド	項目1	項目2	...	因子1の因子得点	因子2の因子得点	因子3の因子得点	因子4の因子得点
A	X	1	1					
B	X	0	1					
C	X	1	0					
D	X	0	1					
E	X	1	0					
F	X	0	1					
G	X	1	1					

AさんのブランドAに対する項目1の回答
= 因子1負荷量×因子1の因子得点+因子2負荷量×因子2の因子得点
+ 因子3負荷量×因子3の因子得点+因子4負荷量×因子4の因子得点
となるような得点が因子得点1~4に計算される

Copyright (C) Nomura Research Institute, Ltd. All rights reserved. NRI 26

図 24 因子分析の概要

(4) 第4回研修会 「音楽×AI×データサイエンス」

最新の音楽に関する様々なデータサイエンス研究の事例発表であった。

高校生にとって、データ収集は非常に困難なプロセスであるが、本研修ではオープンデータで採取できる音楽関連データセットの紹介があった(図25)。また、音楽には様々な観点があるため、どのデータをどのように扱うかどうかによって、研究の方向性が変わってくる(図26)。

データ(D1収集)

【指導の視点①】

データセットの紹介

- AIやデータサイエンスにはデータが必要
 - 音楽に関連するデータセットが公開されている
 - 多くは研究目的・非商用に限りダウンロード可能
 - <https://www.audiocontentanalysis.org/data-sets/>
- データセットの紹介
 - MIDI(楽譜)や和音のデータセット
 - Lakh MIDI Dataset, McGill Billboard
 - ダンス動作のデータセット
 - AIST Dance Database, AIST++ Dataset
 - 音楽音響信号のデータセット
 - MagnaTagATune, MAESTRO

データセット等のURL一覧



- 音楽関連データセット一覧
 - <https://www.audiocontentanalysis.org/data-sets/>
- Lakh MIDI Dataset
 - <https://colinraffel.com/projects/lmd/#get>
- McGill Billboard annotations
 - https://ddmal.music.mcgill.ca/research/The_McGill_Billboard_Project
- AIST Dance Database
 - <https://aistdancedb.ongaaccel.jp/>
- AIST++ Dataset
 - https://google.github.io/aistplusplus_dataset/index.html
- The MagnaTagATune Dataset
 - <https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset>
- MAESTRO dataset
 - <https://magenta.tensorflow.org/datasets/maestro#v200>
- librosa
 - <https://librosa.org/doc/latest/index.html>
- pretty_midi
 - <https://craffel.github.io/pretty-midi/#pretty-midi-instrument>

図25 音楽に関連するデータセット一覧

データ(D1収集)

【指導の視点②】

メロディ作曲における制約条件

歌詞 例: はし  <p>歌詞のアクセントと音高上下の関係</p>	和声進行  <p>和声: GM CM 旋律各音の和声への適合</p>	リズム  <p>1曲中で一貫したリズムを生成するための制約</p>
伴奏  <p>連続5度・8度などの作曲法の禁制</p>	音域  <p>歌にくい音域や頻出する音域の設定</p>	跳躍  <p>歌にくい跳躍や頻出する音域の設定</p>

図26 音楽分析する際の観点の一例

可視化する観点によって、特徴が見出せるかどうか変わってくるため、可視化を繰り返しながら、結論につながる特徴を見抜いていく（図 27）。また、図 28 では、波形が 0 をまたぐ回数や信号の長さによって測定するという分析方法の一例が紹介された。

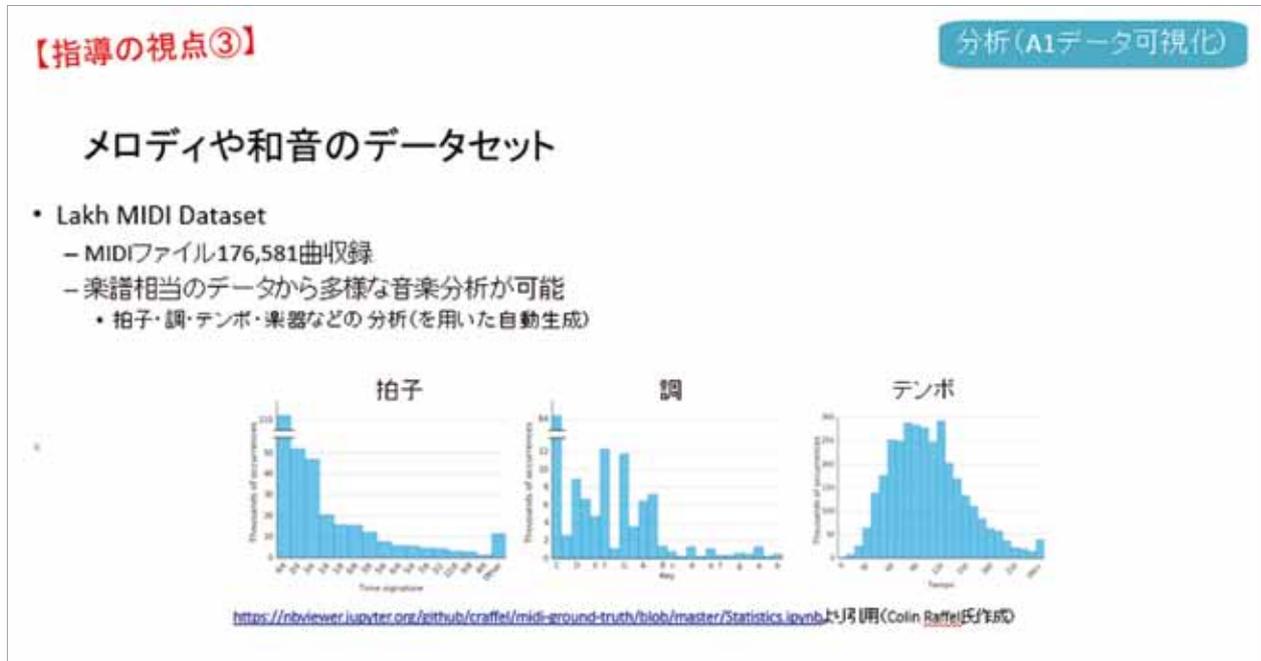


図 27 音楽分析する際の観点の一例

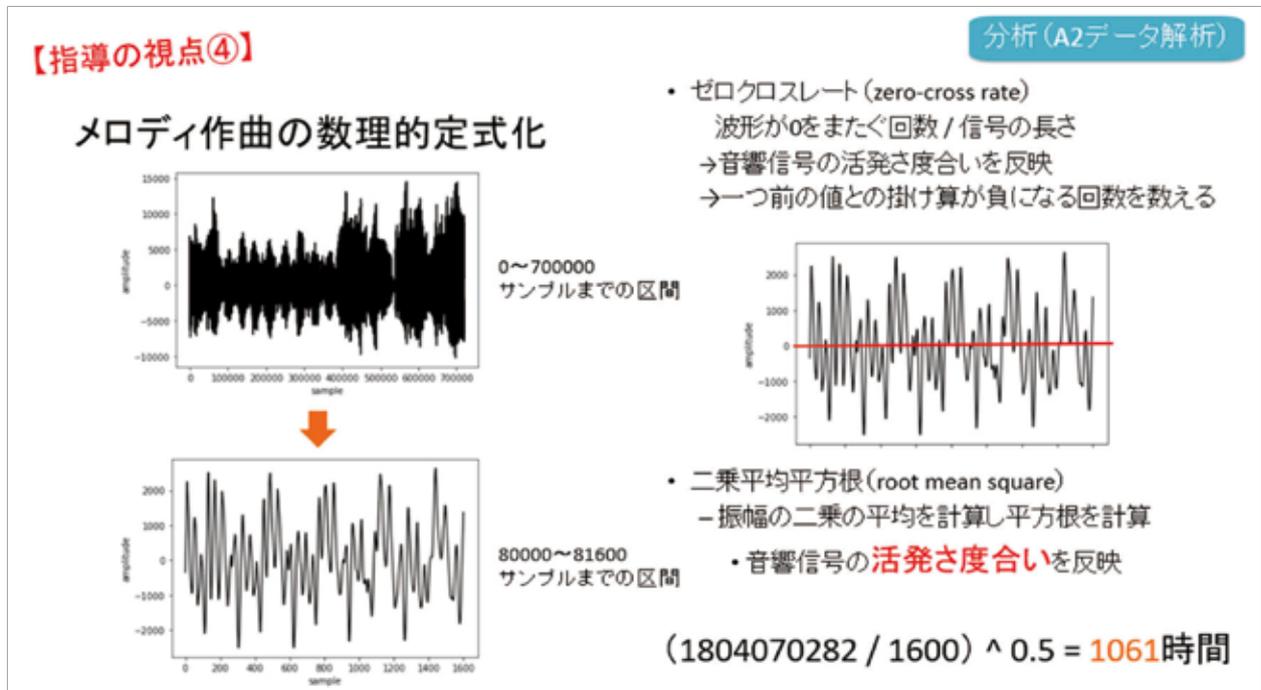


図 28 メロディ作曲の数理的定式化の概要

回帰分析法やニューラルネットワークの分析手法の紹介があった（図 29， 30）。

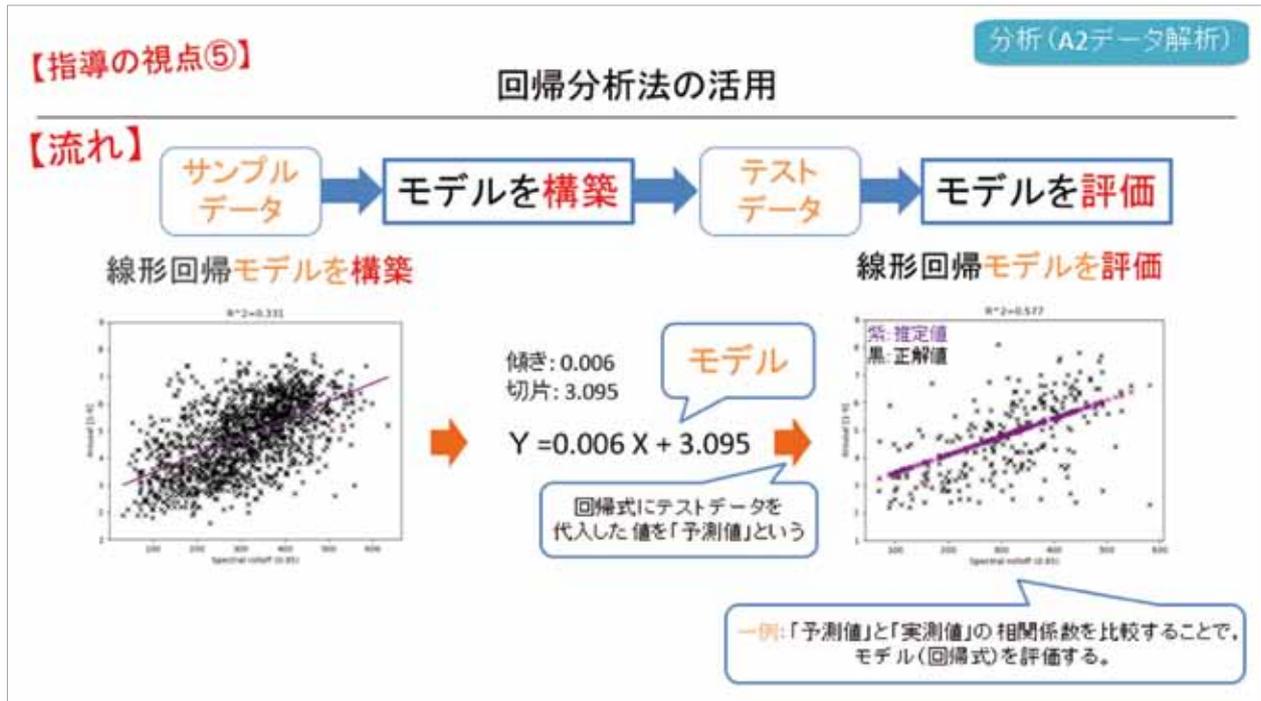


図 29 回帰分析法の概要

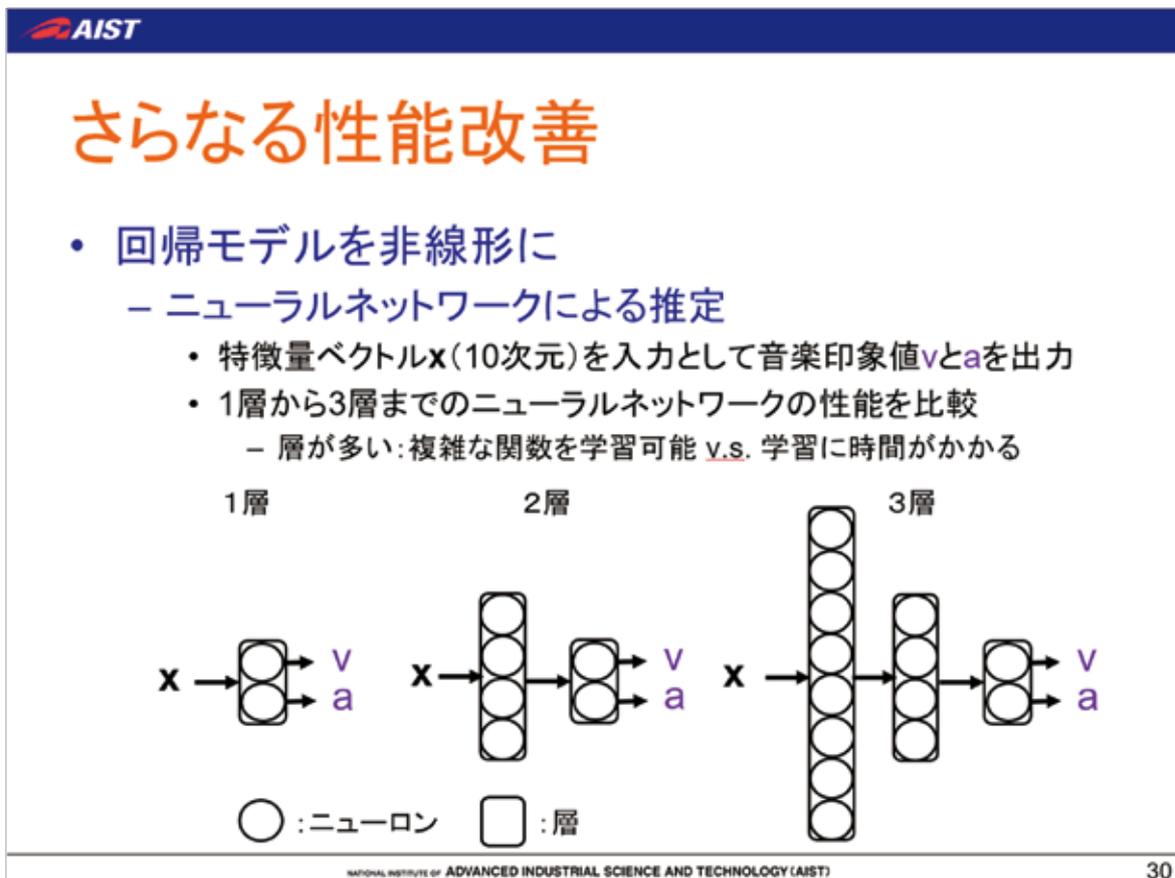


図 30 ニューラルネットワークによる推定の概要

(5) 第5回研修会 「スポーツ×データサイエンス」

フィギュアスケートを題材としたデータサイエンスの研究事例発表であった。

どのような要素は得点が高いのか、どのように改善すれば得点が高くなるのかという分析は大切である。正確にスポーツの特性をつかみ、得点につながる要素を可視化している (図 31, 32)。

分析事例のお品書き

-  1. フィギュアスケートってどんな競技？
映像分析ツールを使ったデータ取得と分析一技のつなぎとはー
(基本：記述統計)
-  2. トラッキングデータから見る出来栄の良いジャンプとは？
(応用：分散分析・多重比較検定)
-  3. 出来栄点に相関がある新たなジャンプの評価軸を探そう！
(発展：主成分回帰分析)

HP JAPAN SPORT COUNCIL SC JAPAN SPORT COUNCIL 50

図 31 分析の全体像

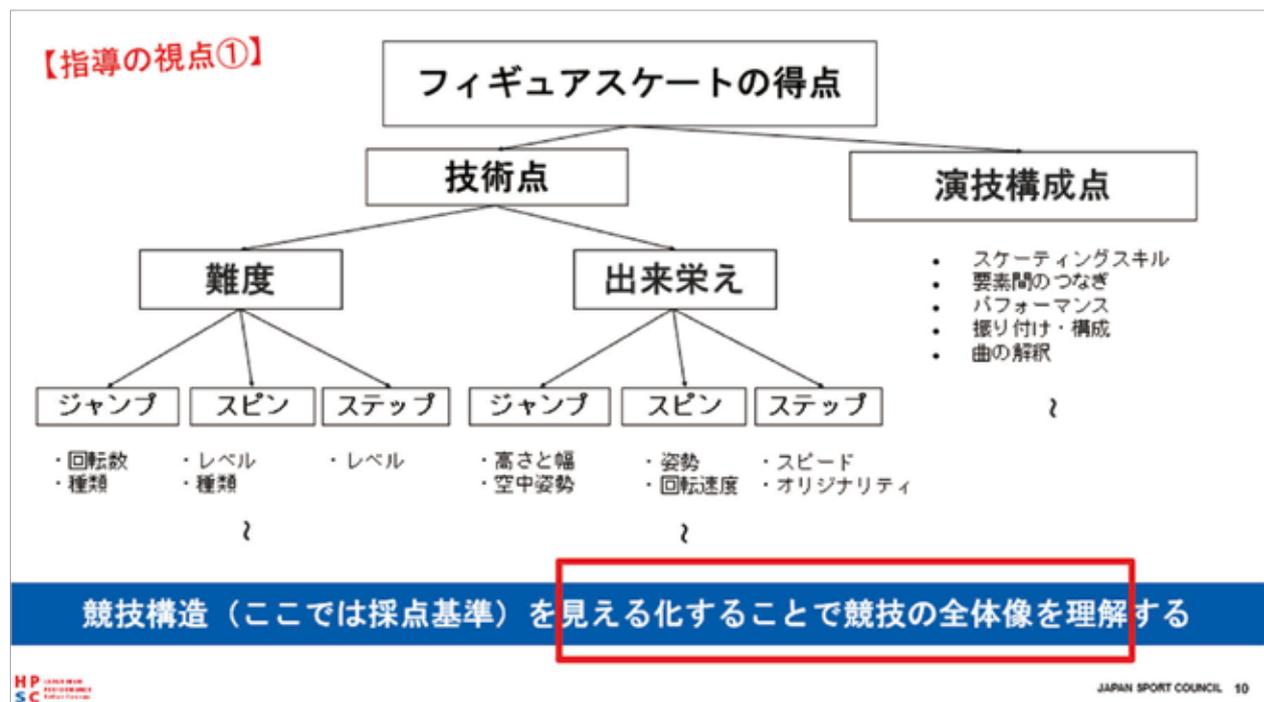


図 32 採点基準の見える化

研究において言葉の定義を具体的に示すことは重要である。採点基準にある「良い」って何？と疑問を抱く視点を持ち、「良い」を具体的に解明する研究事例である(図33)。また、ジャンプであれば、どのような視点で、どのように計測するのかを定めることによって、データ化できる。データ化する際の基準を定めることで、分析のプロセスの充実につながっていく(図34)。

ジャンプの出来栄え点の加点方面のガイドライン

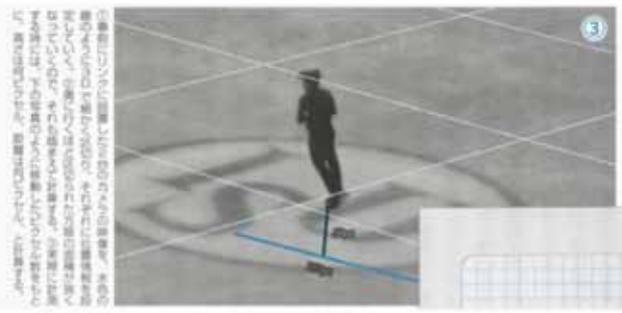
1	高さおよび距離が 非常に良い
2	踏切および着氷が 良い
3	開始から終了まで無駄な力が全く無い
4	ジャンプの前にステップ、予想外または創造的な入り方
5	踏切から着氷までの身体の姿勢が非常に良い
6	要素が音楽に合っている

「良い」って何？

HP SC JAPAN SPORT COUNCIL 102

図33 「ジャンプ」の出来栄え点に関するガイドライン

【指導の視点②】
研究方法：データ取得 ー説明変数ー



IceScope

- ・幅(m)
離氷・着氷フレーム間の水平距離
- ・高さ(m)
離氷～着氷フレーム間の垂直成分の最大値
- ・離着氷フレーム
開発会社担当者が目視で判断
ブレードのつま先が基準点

HP SC JAPAN SPORT COUNCIL 12

図34 「ジャンプ」をデータ化する観点・基準

出来栄点によって分類し、平均値の差の検定を行うなど、データを分類・比較したり（図 35）、主成分分析法を用いて、主成分の得点でグループを作成して、比較・解析したりなど、特徴を見出すまで試行錯誤を行い、データの整理整形、可視化、解析を繰り返すことが重要である。また、クラスターに分類したあと、その特徴を考察することが研究の醍醐味である（図 36）。

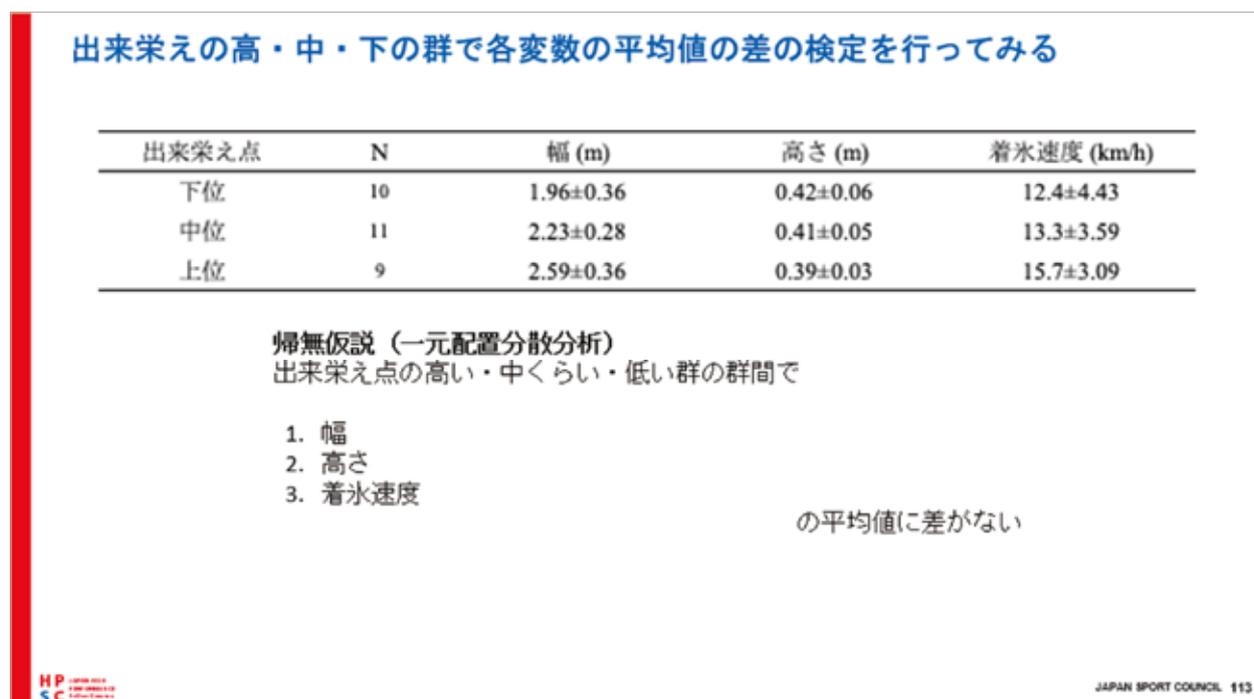


図 35 出来栄点の群ごとの検定の実施

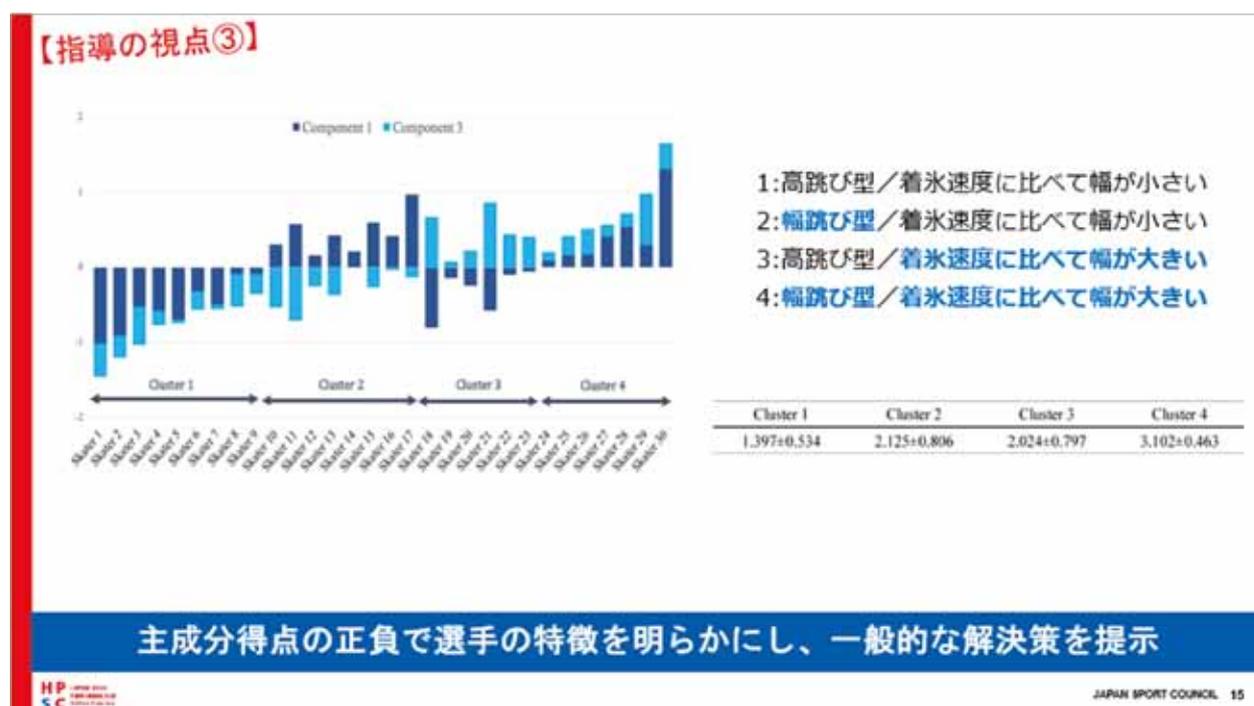


図 36 主成分の寄与率でクラスターに分類し、導き出した結論

(6) 第6回研修会 「情報教育×データサイエンス」

令和4年度から実施される「情報Ⅰ」「情報Ⅱ」の学習指導要領の要点，データサイエンスに関わる小・中・高のつながり（図37），他教科との連携など，データサイエンスを学校教育として体系的に取り組む時代が到来した（図38）。

小学校からの学習の積み上げ			
	情報デザイン	プログラミング	統計に関連した学び
情報Ⅱ	情報デザインを生かしたコンテンツ作成	情報システムのプログラミング	データサイエンス ※数学Bと連携
情報Ⅰ	情報デザインの方法と考え方 問題を発見・解決する手段として活用	問題解決のためのプログラミング コンピュータの仕組み モデル化・シミュレーション	データの活用 ※数学Ⅰと連携
中学校	技術・家庭科など 中学校の各教科等	問題解決のための簡単なプログラミング 計測・制御 ネットワーク&双方向	簡単な統計
小学校	国語，図画工作など 小学校の各教科等	教科の中で体験するプログラミング 仕組みを知り，活用して可能性を広げる	統計的考え方

図37 情報における学習指導要領における小中高のつながり

「情報Ⅰ」はデータサイエンスの基礎

- 最初は中学校までの数学を使って問題の発見・解決
- プログラミングではシミュレーションで確率なども入ってくる
「数学A」の「場合の数と確率」との連携は学習指導要領解説に記載
- 「数学Ⅰ」の「データ分析」と連携
- 実際にデータを扱うために必要になることは，「数学Ⅰ」になくても「情報Ⅰ」でやる
 - 整理されていないデータ
 - 欠損値
 - 尺度水準

図38 情報Ⅰのデータサイエンスに関わる概要

「情報Ⅰ」においては、問題解決に関するプロセスやデータの扱いなど、研究に関わる内容が多く含まれており、数学、総合的な探究の時間、課題研究などと連携することでより効率的な教育活動となっていく（図 39, 40）。

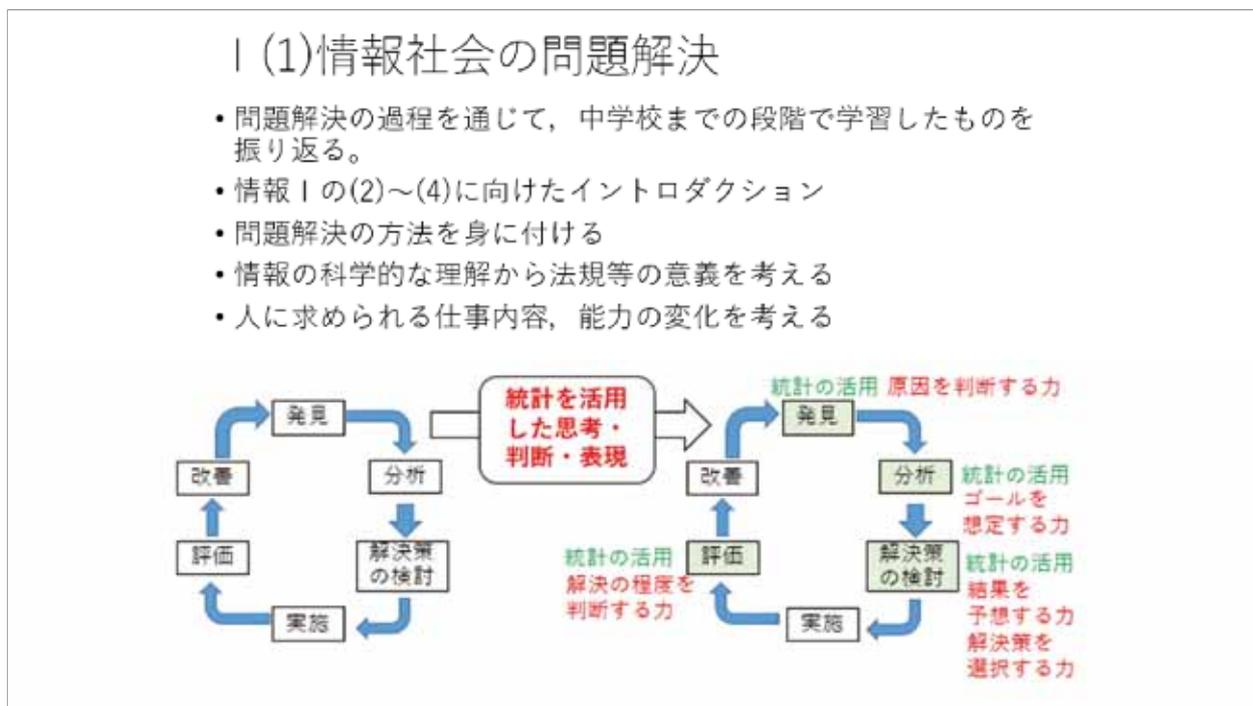


図 39 情報社会の問題解決の概要

情報Ⅰ(4) データの扱い

	「社会と情報」	「情報の科学」	「情報Ⅰ」
統計	数学と連携して 平均値、中央値 などの基本的統計値を扱う		分散、標準偏差、相関係数などの統計指標、散布図、仮説検定の考え方、交絡因子なども扱う
分析	主にグラフ化などを行い、データの傾向をつかむ		クロス集計、仮説検定、重回帰分析、これらを通じたデータの可視化、現象のモデル化と予測
量的データ	主に表形式で整理された数値を中心に扱う		量的データの記載あり。表形式で整理されていないものも扱う
質的データ	質的データの記載なし テキストマイニングの例あり		質的データの記載あり テキストマイニングの例あり
扱うデータ	整理されたデータを扱う		実験値などの整理されていないデータも扱い、外れ値、欠損値などの処理も学ぶ
尺度	-		名義、順序、間隔、比例など尺度水準の違いを扱う
データベース	「情報の科学」のみで扱う		情報を収集・蓄積・提供する方法として全員が学ぶ

中学校数学科「Dデータの活用」、高校「数学Ⅰ」の(4)「データ分析」と連携
 赤字＝数学科で学び情報科で活用 赤空＝情報科のみで活用

図 40 「社会と情報」から「情報Ⅰ」の変遷

「情報Ⅱ」においては、データの活用の分野では専門的な統計手法を学ぶことができ、数学Bの内容も含まれていることから、数学と連携した取組が効果的である。また、産業界を中心にデジタル関連部活のニーズを踏まえた支援が始まる時代の到来が示された（図41、42）。

情報Ⅰ(4) 情報Ⅱ(3) データの活用		
	「情報Ⅰ」	「情報Ⅱ」
統計	分散、標準偏差、相関係数などの統計指標、散布図、検定の考え方、交絡因子なども扱う	統計的な推測（標本調査、母集団の特徴や傾向）、仮説検定の方法などを扱う
分析	クロス集計、仮説検定、単回帰分析、これらを通じたデータの可視化、現象のモデル化と予測	重回帰分析、分類、クラスターリング、これらを通じた可視化、現象のモデル化と予測及びモデルの評価、機械学習
量的データ	量的データの記載あり。表形式で整理されていないものも扱う	多機かつ大量のデータを扱い、バイアスなどデータの信頼性にかかわることにも配慮する特に記載なし
質的データ	質的データの記載あり。テキストマイニングの例あり。	
扱うデータ	実験値などの整理されていないデータも扱い、外れ値、欠損値などの処理も学ぶ	
尺度	名義、順序、間隔、比例など尺度水準の違いを扱う	
データベース	情報を収集・蓄積・提供する方法として全員が学ぶ	データの整形などで、データを扱うプログラミングにも触れる

中学校数学「Dデータの活用」 高校「数学B」の(2)「統計的な推測」
 赤字＝数学科で学び情報科で活用 赤字＝情報科のみで活用

図41 「情報Ⅰ」「情報Ⅱ」におけるデータの活用

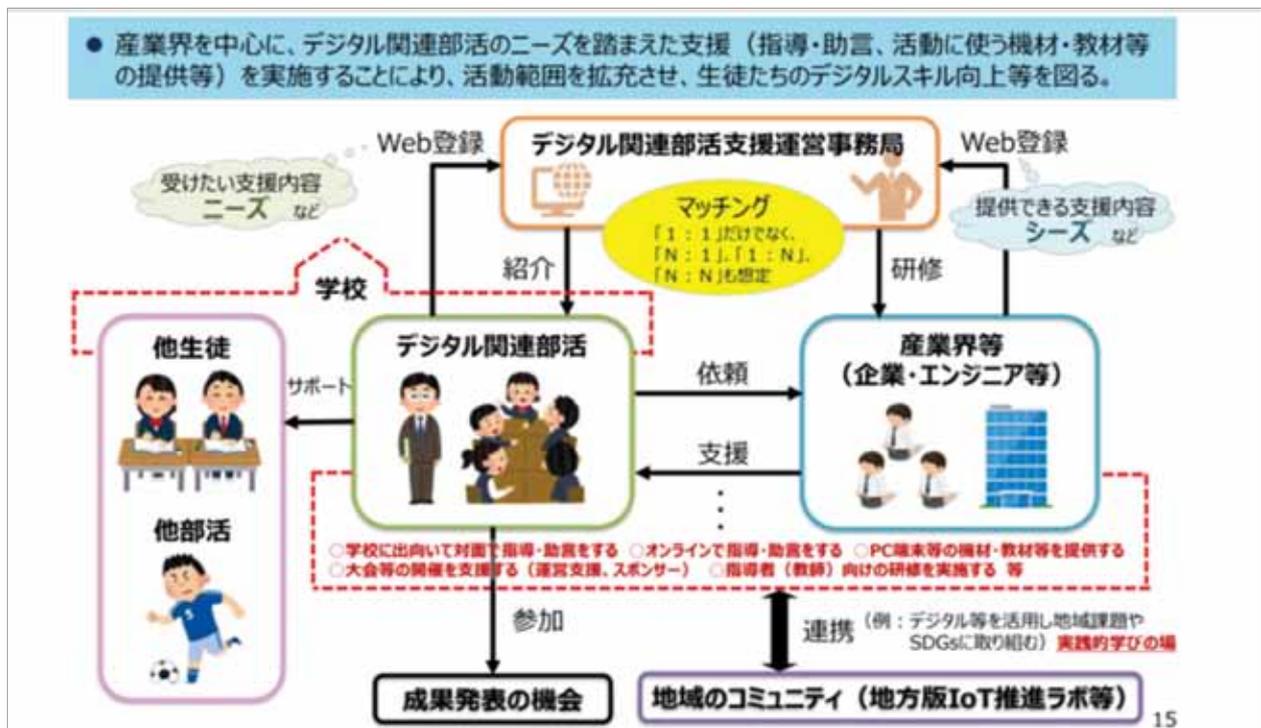


図42 想定される学校とデジタル関連部活を支援する社会体制

文部科学省では、高等学校情報科「情報Ⅰ」「情報Ⅱ」における教員研修用教材を作成し、公開されている。非常にわかりやすく、具体的な事例を含んだ内容となっており、情報科だけでなく、探究活動、課題研究を指導する教員が参考にできるものである（図43、44）。

高等学校情報科「情報Ⅱ」教員研修用教材（本編）



文部科学省 MINISTRY OF EDUCATION, CULTURE, SPORTS, SCIENCE AND TECHNOLOGY (MEXT)

[トップページ](#) > [English](#)

会見・報道・お知らせ
政策・審議会
白書・統計・出版物

トップ > 教育 > 小学校、中学校、高等学校 > 教育の情報化の推進 > 高等学校情報科(各学科)に共通する教材 > 高等学校情報科「情報Ⅱ」教員研修用教材(本編)

● 高等学校情報科「情報Ⅱ」教員研修用教材(本編)

- ① 表紙はじめに、目次、序章 (PDF 2.8MB)
- ② 第1章 情報社会の進展と情報技術 (PDF 8.6MB)
- ③ 第2章 コミュニケーションとコンテンツ (PDF 9.7MB)
- ④ 第3章 情報とデータサイエンス 前半 (PDF 8.9MB)
- ⑤ 第3章 情報とデータサイエンス 後半 (PDF 7.6MB)
- ⑥ 第4章 情報システムとプログラミング (PDF 4.2MB)
- ⑦ 第5章 情報と情報技術を活用した問題発見・解決の探究 巻末 (PDF 4.1MB)

https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_00742.html

図 43 文部科学省高等学校情報科「情報Ⅱ」教員研修用教材

【指導の視点①】 高等学校情報科「情報Ⅱ」教員研修用教材（本編）

1 課題発見とデータに基づく問題解決：予測と制御

問題解決をデータに基づいて行う場合、まず解くべき課題の発見、把握が必要である。課題発見とは、例えば、海の環境やごみ置き場の清掃状況、地域の商店街の様子や所属するスポーツ部の成績や試験結果など、身の回りの現象に対して、現実があるべき姿(理想)の状態ではないことを知覚し、現実と理想のそれぞれの状態の明確化とそのギャップが解くべき課題である

ことを具体的に示す力をいう。次に、その課題をデータと分析で解決する科学的な問題解決のフレームに基づきしむためには、理想や現実の状態を示す客観的なデータ指標Y(目的変数、ターゲット変数、教師変数、予測変数、被説明変数)を具体的に定め、その値の変化や変動に何が影響するのかわかると、5W1Hやその発生に至るプロセス要因

勝つためにはゴールを決める事が大事だよ

どんなプレーがゴールにつながるだろう？

図解1 特性要因図による要因の洗い出し

図 44 「情報Ⅱ」教員研修用教材の抜粋

謝辞

講師の先生方には、最新のデータサイエンスの研究事例、実践事例を学ぶ機会をいただき、心より感謝申し上げます。

研修会に参加された全国の高等学校教員の皆様方におかれましては、公務ご多忙の中、参加いただき、また貴重なご意見をいただきました。

最後に、本研修会は、令和3年度SSH交流事業による支援によって実施しましたが、皆様方のご協力により、充実した研修会を開催できたことを感謝申し上げます。

今後とも、兵庫県立姫路西高等学校の教育活動についてご指導ご協力をお願い申し上げます。

兵庫県立姫路西高等学校

発刊日 令和4年3月

兵庫県立姫路西高等学校

〒670-0877 姫路市北八代二丁目1番33号

TEL (079) 281-6621 (代)

FAX (079) 281-6623

Home Page <https://www.hyogo-c.ed.jp/~himenisi-hs>

E-Mail himenisi-hs-ssh1@hyogo-c.ed.jp

