

高等学校 第1学年 情報科 学習指導案

SSH 学校設定科目「データサイエンス入門」(1単位)

期 日：2021年11月12日(金)

時 間：第6校時(14:50~15:40)

対 象：第1学年D組生命科学コース12名

学校名：清心女子高等学校

場 所：高校普通教室棟2階 高1D教室

授業者：坂部 高平

1 単元名

「データの活用」

2 単元の目標

- (1) データを表現、蓄積するための表し方と、データを収集、整理、分析する方法について理解し、技能を身に付ける。
- (2) データの収集、整理、分析及び結果の表現の方法を適切に選択し、実行し、評価し改善する。
- (3) 問題の発見や解決にデータを活用するために、適切なデータを選択や、分析の仕方、解釈の仕方について、粘り強く取り組み、試行錯誤を通じて改善しようとする。

3 単元の評価規準

知識・技能	思考・判断・表現	主体的に学習に取り組む態度
<p>データを問題の発見・解決に活用するために、データを収集、整理、分析する一連のデータ処理の流れ及び、データの特徴を表す指標と、その評価について理解している。</p> <p>「質的データ」と「量的データ」の性質や扱い方の違いを理解する。データの内容や形式を踏まえて処理する技能を身に付けている。</p> <p>データに含まれる欠損値や外れ値の扱いやデータを整理、変換する必要性を理解する。基礎的な分析及び可視化の方法を理解するとともにその技能を身に付けている。</p>	<p>必要なデータの収集について、選択、判断し、それに応じて適切なデータの整理や変換の方法を判断することができる。また、分析の目的に応じた方法を選択、処理し、その結果について多面的にアプローチし、可視化を行うことにより、データに含まれる傾向を見いだすことができる。</p> <p>データの傾向に関して評価するための客観的な指標を選択して利用し、自身の考えを基にした適切な解釈を行うことができる。</p>	<p>問題の発見・解決にデータを活用するために、適切なデータを選択や、分析の仕方、解釈の仕方について試行錯誤し、粘り強く取り組んで改善しようとしている。</p>

4 指導観

(1) 単元観

これからの先の社会である Society 5.0 においては、モノのインターネット (Internet of Things=IoT) がさらに進む。現実社会のあらゆるものがセンサーなどでインターネットとつながるので、その情報量は飛躍的に増大し、その処理は人の能力を超えたものになるものと考えられる。そこでその処理のために人工知能 (AI) の活用が進む。それら社会を支えるしくみについて、その構成員が無知でいてよいはずはない。情報が経済的な発展の道具から、環境面や人の暮らしを含めた社会基盤を支える道具になる世界での活躍が期待される世代に向けて、統計に関する内容を中心に扱う中で、データの利活用に関する科学的な知識や理解を促す。また、問題の発見や解決に向けて適切な手法でデータを手に入れ、それを効果的に用いる技能と、それをさらに粘り強く改善していく姿勢を養うことを目指す。

(2) 教材観

SSH 学校設定科目であり、教科書の内容を扱う科目ではないため、情報科の科目内容の一部を扱うが、教材はオリジナルで作成したものを使用する。本単元はデータの活用がテーマであるため、統計に関する内容が中心である。ICT 機器やそれを利用するアプリケーションはあくまでも道具であり、今手にしている道具を効果的に利用していくための基本的な知識・技能と、それを自分自身で必要に応じて発展させるための論理的な思考を身につけることを意識している。今年度の高校 1 年生からは、全員が ICT 機器として同一機種種の Chromebook を手にしており、いつでもどこでもインターネットに接続された同じ道具を利用できる環境が整った。データサイエンスについて考え学んでいく上で、表計算は避けて通ることのできないものと考え、Google スプレッドシートを利用した実習を中心に授業を展開する。教材のやりとりについては Google Classroom を利用し、オンラインでの小テストも授業内容のまとめりごとにも実施して Classroom で返却したり、ファイルによる課題提出を Classroom で回収したりすることで、学習の過程が自動的にオンライン上に蓄積された状態で散逸せず、課題の取り組み状況の確認もオンライン上でおこなっている。

データの処理に取り組むにあたり、扱うデータについては模式的なものを準備し、まずはその処理のしくみの根本的な理解を促す。その次の段階では公的な実際のオープンデータを用いた作業も取り入れ、データサイエンス、特に統計のことを身近に感じさせ、自分たちも扱うことができるものであり、有用な道具・手法であるという意識を持たせたい。

(3) 生徒観

全員が ICT 機器として同一機種種の Chromebook を手にしており、各教科の授業でも利用されていることから、PC の操作に不安がある生徒は皆無である。特に、この授業対象となっている生命科学コースの生徒は、理系の科目や進路に興味をもつ生徒が多くを占めていることもあり、PC 利用にアレルギー反応を示すような生徒は見られない。単なるメモやまとめ書き、整理する枠として利用する以上に、本格的に Google スプレッドシートを表計算の道具として利用したことがない生徒もいるだろうが、表計算という枠組みのアプリケーションに触れたことのない生徒はいない。Google Classroom も様々な授業で利用されていることから、利用方法に戸惑うことはない。

少人数での授業になっていることも影響し、本科目の授業時間内においては、各自が静かに作業に取り組む印象である。グループワークを一部に組み込んで、お互いで情報交換をし、教え合い、助け合う場面を作り、協働による学習意欲の向上と技能習得の促進を図りたい。

5 年間指導計画における位置付け

学年	第1学年 (1単位)	第2学年	第3学年
単元名	①情報社会の問題 ②表計算の利用 ③データの活用 (←本時の単元) ④プログラミングとシミュレーション		
主な内容	①セキュリティの知識と意識づけ, ネット上の人権意識の涵養 ②表計算の道具としての機能を知り, 利用するための技能の習得 ③データの種類や分析の手法を知り, 統計のための基礎技能の習得 ④Pythonの基礎的な技能の習得とその応用		

6 単元の指導計画と評価計画

指導計画	第一次	データ種類と特徴.....	1時間
	第二次	データの特徴と統計量.....	2時間
	第三次	データ間の関係性.....	2時間
	第四次	データの傾向による分類.....	3時間 (第2時 本時)
	第五次	データからの予測.....	3時間
	第六次	データの倫理.....	1時間

[第四次 指導計画]

時	★目標 ○学習内容 ・学習活動	■評価規準 (評価方法)
第1時	★クラスター分析について知り, ウォード法による具体的な分析手順について確認する。	■知識・技能 スプレッドシート上でウォード法の計算として分散の和を求めることができる。
	○クラスター分析の概要を知る。 ○ウォード法の分析手順を知る。 ・模式的なデータを用いて, Google スプレッドシートでウォード法の計算について実習を通じて体感する。	
第2時 (本時)	★ウォード法によるクラスター分析を実習で体験し, 分析方法を身に付ける。	■知識・技能 樹形図を書くことができる。 ■思考・判断・表現 理由をもって2変数を選んでいる。 ■主体的に学習に取り組む態度 協働的に粘り強くウォード法に取り組むことができている。
	○デンドログラム (樹形図) の意味を知る。 ・政府統計のオープンデータを使ってウォード法によるクラスター分析を試み, デンドログラムを書く。使用する多変量データのうち, 分析に用いる2変数はクラスターに分かれそうな組合せを自分たちで選ぶ。	
第3時	★クラスター分析の結果を確認し, 散布図と比較してその妥当性を協議する。	■思考・判断・表現 分析において妥当な理由をもってクラスターの数を決定できている。 ■主体的に学習に取り組む態度 グループ協議において積極的に議論に参加できている。
	○クラスター分析の妥当性を考える。 ・前時にウォード法で用いた2変数について Google スプレッドシートで散布図を描かせ, 前時のデンドログラムと見比べて, クラスター分析の妥当性を協議する。	

7 指導に当たって

- (1) 本単元第四次の目的はクラスター分析の方法の一つを知ることであり、ウォード法を完璧に間違いなく行うことではない。ウォード法以外にも階層的クラスタリングの方法があり、さらに非階層的な方法もあるが、その細かな手順や原理については触れず、散布図を見ながら視覚的に何となくできそうなクラスタリングが、統計量の一つである分散を用いる計算のできる点に着目させたい。そのため、クラスタリングしやすくするためのデータの加工やクラスタリングの際に必要な組合せパターンの列記についてもここでは詳細には触れず、分析に使用するデータセットをもとにあらかじめ準備しておく。
- (2) 課題解決につながる思考法としての論理的思考・批判的思考・水平思考の3つを養う授業展開を目指している。それらの思考を促す場面の設定を組み込んでいく。本単元第四次においては、生徒個人でクラスター分析をおこなう2変数の組合せをいくつか考える場面において、その変数を選ぶ理由を意識させることで論理的思考（ロジカルシンキング）を促す。そして、実際にウォード法を実践する前に少人数のグループワークで2変数の組合せを1つに絞り込む場面を設定する中で、批判的思考（クリティカルシンキング）を促す。また、変数となるデータの項目を多めに準備することで、その組合せパターンはかなり多岐に渡る。最終的に各グループがどの変数を使ってクラスター分析をしたのか、その結果どのようにクラスタリングされてどのような原因が考えられるかを全体で共有することで、自分の中になかった考え方を発見させ、水平思考（ラテラルシンキング）の育成につなげたい。

8 本時案（第四次 全3時間中の2時間目）

(1) 本時の目標

- 実際の非調査データを使ったウォード法の実践を通して、データ解析の準備→要約→分類→予測のうちの「分類」の工程として、クラスター分析の方法を知る。
- お互いの分析準備や結果を共有することで、同じデータを見ている、さまざまな視点からデータを見ることができると気づく。また、それによって、お互いが補い合うこともできる可能性に気づく。

(2) 本時の展開

時間	○学習内容 ・学習活動	・指導上の留意事項	■評価規準（評価方法）
導入 (8分)	<ul style="list-style-type: none"> ○本時は、クラスター分析を実践することを知る。 ・クラスター分析に関するオンラインでの小テストによって復習する。 	<ul style="list-style-type: none"> ・ウォード法についてその方法を思い出させる。 	<ul style="list-style-type: none"> ■知識・技能 ウォード法の手順を理解している。
展開 (40分)	<ul style="list-style-type: none"> ○デンドログラム（樹形図）を知る。 ・プリントを配布し、デンドログラムの書き方を確認する。 ○分析に用いる2変数を決定する。 ・数人グループで理由を含め話し合っ て変数の組合せを1つ決めさせる。 ○ウォード法を実践する。 ・計算用のスプレッドシートを配信し、 グループで協働してデンドログラム を書きながら分析する。 	<ul style="list-style-type: none"> ・デンドログラムを例示し ウォード法の進め方イ メージをつかませる。 ・机間指導で、論理的、批 判的に話し合っている かを確認する。 ・以下の点を見ながら机 間指導をする。 ①グループ内で協力 ②手順の正しさ 	<ul style="list-style-type: none"> ■思考・判断・表現 論理的・批判的な議論で意見を まとめることができている。 ■主体的に学習に取り組む態度 グループ内で積極的に協力して 作業できている。 ■知識・技能 ウォード法が実践できている。

まとめ (2分)	○次回の内容を知る。 ・次回は散布図を作成し、今回の結果と散布図を見ながらクラスタリングの妥当性を検討することを伝える。	
-------------	---	--

(3) 板書計画

今回授業配布プリント→

データサイエンス入門 授業プリント

データの分類 (2) 【クラスター分析】

分類結果を可視化する

前回登場したクラスター分析を可視化する方法を見てみましょう。

● (デンドログラム) …日本語では (樹形図) とも呼ばれる。階層的分類の結果を可視化するグラフである。トーナメント表のような形になる。

→前回のウォード法練習の分析をグラフにすると、次のようになる。この場合のクラスターに分けるとしたら①～③の箇所が考えられるが、右の散布図とも比較してみると、③の段階での2つ (点線円で囲んだ2つ) が妥当と考えられる。

→政府統計から取り出した非調査データを使った分析を進めましょう！

【練習 2-1】 スプレッドシート (Classroom で配付する「分類練習 2」) にウォード法のためのデータ (データ項目の組合せと分散の表) があります。近くの席の人とグループを作って相談し、練習 1-3 で考え「分類練習 1」のシート 1 に記入した 2 つのデータ項目の組合せから、分析する組合せをグループで 1 つ選びましょう。そして、表の右端に 2 つのデータ項目の分散の合計値を計算させて、ウォード法でデンドログラムにしてみましょう。(このプリントの裏面を使う。)

高 ID ____ 番 名前 ____

政府統計 _____ と _____ のクラスター分析デンドログラム

	1	2	3	4	5	6	7	8	9	10
都道府県										
○●県										
○△県										
・・・県										
...										
...										
...										
...										
...										
...										
...										
...										
...										

付する「分類練習 2」) を使って、練習 8-1 し、練習 2-1 で作成したデンドログラムと比較し、グループで話し合ってみましょう。

グラフ上のプロットにポインターを重ねたときに、点の色が確認できます。

データサイエンス入門 検業プリント

データの分類 (1) 【量的データの分類法】

似たもの同士を分類する

多くの変数 (量的データ) を含むデータセットでは、さまざまな変数を組み合わせることでそれらを分析すると、さらなる特徴や傾向が見えてくる場合があります。そのようなときのデータ解析の方法は「多変量データ解析」と呼ばれます。そのうちの一つとして、類似したデータを直接分類する方法を見てみます。

- (クラスタ) 分析…多くの変数 (データ) から

見たときに、それらをまとめて、似たパターンを示す

データポイント进行分类するデータ解析の方法。

2種類の变数で散布図にすると、データポイントが似たもの同士の「群れ」を作ることがある。この「群れ」のことを (クラスタ) と呼ぶ。

→右の散布図だと点線で囲んだ3つ。

でも「？」はどこに属すべきか判断が難しい。

→政府統計から取り出した非調査データを使って、実際に分析を進めていきます。

【利用する2015年産政府統計データ】

都道府県：北海道、青森県、群馬県、東京都、新潟県、愛知県、大阪府、岡山県、香

川県、福岡県、沖縄県の11都道府県

データ (変数)：総人口 [人] 総人口(男) [人] 総人口(女) [人]

15歳未満人口 [人] 65歳以上人口 [人]

未婚人口(15歳以上X男) [人] 未婚人口(15歳以上X女) [人]

出生数 [人] 総面積 [km²] 可住地面積 [km²]

1人当たり県民所得 [千円] 高等学校数 [校] 大学数 [校]

完全失業者数 [人] 第1次産業就業者数 [人]

第2次産業就業者数 [人] 第3次産業就業者数 [人]

牛乳ストローク送信者数 [人] 電灯使用電力 [百万kWh]

1人1日当たりのゴミ排出量 [g/人・日] 刑法犯認知件数 [件]

【練習1-1】クラスタ分析の準備をします。まずはそれぞれの変数 (データ) の特徴を統計値によって確認します。スプレッドシート (**Classroom**) で配付する「分類練習1」の

シート1) についていくつかの都道府県の21種類のデータをあります。各変数 (項目) の統計値を表の下に加えます。今回加えてみる統計値は、(平均) 値 **Average**、

(中央) 値 **Median** (標準偏差) **Standard Deviation**、**STDEV.P**、変動係数 **Coefficient of Variation (CV = Average)** の4つです。ただし、「全国」を計算に含めてはダメです。

- (クラウド Word) 法…クラスタ分析の計算方法の一つ。個々のデータポイントを段階ごとに分類していく方法。これは (階層的) 分類と呼ばれるものの一つ。計算量が多くなるのはデメリットだが、パワースクワ分類できる点が優れている。

【手順】各クラスタ1間 (最初は各データポイント) の各変数の (分散) **Variance** (開散) は **VAR.P** の合計値を比較し、併合後のクラスタと併合前のそれぞれのクラスタの和が最小のものを組み合わせて1つのクラスタにする。

これを繰り返してクラスタを段階的にまとめていき、クラスタが1つだけになったら分析終了。

【具体的な計算方法例】次のようなデータの場合

	変数 X	変数 Y
対象 A	4	5
対象 B	2	7
対象 C	8	9
対象 D	4	1
AB 分散	1	1
BC 分散	9	4
AC 分散	4	1
...
AIRC 分散	62	27
...

Yの値とXの値

AとBの分散の計=2 BとCの分散の計=19

AとCの分散の計=5 AとDの分散の計=4

BとDの分散の計=10 CとDの分散の計=5

最小はAB間なので、[AB]をクラスタにまとめる。

まとまる可能性は、[AB]とC、[AB]とD、CとDがあるので同様に分散の計を比較する。

AIRCの分散の計=89でABの分散との差は89-2=69 ABCの分散の計=71で差は51

CDの分散の計=5で最小はCD間なので[CD]をクラスタにまとめる。

最後に[AB]と[CD]をまとめて1つになるので分析終了。

【練習1-2】スプレッドシート (**Classroom**) で配付する「分類練習1」のシート2) に上記の

例の表があるので、分散とその計を計算させてみましょう。(分散の計は自動計算するように、あらかじめ計算式が入力されています。)

ヒット分散は分散を計算可能なので、スプレッドシートのヘルプに尋ねてください。

【練習1-3】スプレッドシート (**Classroom**) で配付する「分類練習1」のシート1) で、変数

(データ) 間に地域差があってクラスタがいくつかできそうな2つの変数を項目名と統計値を見ながら選んでみましょう。3組以上選び、クラスタが複数存在するかどうかを判断した理由を簡単に入力しておきましょう。

※次回、これを使って作業するので、次回までに準備しておいてください。