

高等学校「情報科」におけるデータの活用分野で取り扱う教材を作成するための手掛かり

情報教育研修課 主任指導主事兼課長 安本 靖史
指導主事 原口 攻一郎
指導主事 難波 伸也
指導主事 谷崎 壮志

はじめに

平成 30 年告示の高等学校学習指導要領共通教科情報科において、「情報Ⅰ」が必修科目として開設された。以前の共通教科情報科の必修科目は、「社会と情報」か「情報の科学」のどちらかであり、生徒の選択割合は「社会と情報」が約 80%、「情報の科学」が約 20%であった。

「情報Ⅰ」では、「情報社会の問題解決」「コミュニケーションと情報デザイン」「コンピュータとプログラミング」「情報通信ネットワークとデータの活用」の 4 つの内容を指導する。中でも、「情報通信ネットワークとデータの活用」では、データの形式に関する知識、統計的に分析し結果を可視化する技能、データを基に多面的に考える力、分析結果を分かりやすく伝える力等を養う。扱う内容は、データの形式や分析目的に応じた可視化の方法、相関係数などの統計指標、相関関係や因果関係等のデータの関係性、交絡因子、データの関係性を数式の形で表す単回帰分析等である。

今まで、「社会と情報」の指導が主であった教員にとって、特に、「情報通信ネットワークとデータの活用」の「ア(ウ) 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こと」、「イ(ウ) 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、評価し改善すること」の指導に不安を感じているという声を研修時に聞いた。

そこで、本研究では、高等学校で行うデータ活用分野の授業において生徒に育成する力を整理するとともに、授業で使用するデータを作成するための視点を提案する。

1 研究の概要

本研究では、高等学校情報科及び代替科目担当の教員を対象としたアンケートを実施し、本県における教員のデータ活用（統計）に関する理解度と授業で行われている内容、「情報Ⅰ」の「データの活用」を授業で行う際の課題について把握し、分析・考察した。この課題を解決する提案を行う。

2 令和 5 年度の取組

(1) アンケートによる実態調査

共通教科「情報科」を担当する全ての教員（n=172）を対象に、以下の内容についてアンケート形式で質問した。

- ・データ活用（統計）に関する知識
- ・「データの活用」の授業で用いたデータ例や授業内容で困っていること
- ・「情報Ⅰ」全体を通して困っていること

まず、データ活用（統計）に関する知識の、代表値、分散、ヒストグラム、四分位数、外れ値、相関係数、相関係数の判断基準（相関がある・ない）については、ほとんどの教

員が理解していると回答した。しかし、交絡因子や回帰分析については、あまりよく理解していない部分があると回答した教員もいた。今後、当教育研修所の研修においても丁寧に説明していきたい。

授業内容についての質問では、教科書に記載されているデータ以外にも、オープンデータを使ったり身体測定や体力測定等の生徒の記録を利用したりして、授業を行っている教員がいる。一方で、どの様なデータを使ったらよいか悩んでいる教員が半数程度いた。

また、指導に困っている内容として、「データをどのように活用すれば生徒にとってわかりやすくなるのかわからない」、「統計に関する知識について自信がない部分がある」、「演習で使用するデータが見つからない」がよく挙がっていた。

(2) 「データの活用」に関するサイト

実態調査から明らかになった多くの教員の困っている点の一助とするため、「データの活用」に関するサイトを県立教育研修所 Web ページ内に作成することとした。

サイトの内容は、次の2つである。

ア 教材として利用できるデータ例

イ 授業の進め方と学習評価の例

ア 教材として利用できるデータ例

授業で利用する教材は、生徒に何を理解させるかによってデータを用意する。

(ア) データの傾向を見る授業

分布の広がりを示す分散の授業において、オープンデータ等から収集してもよいが、必ずしも目的のデータが見つかるとは限らない、表計算ソフトウェアの関数を利用し、オリジナルのデータを用意することも一つの方法である。

(イ) データの相関を見る授業

相関の授業において、何を最大（最小）にするのか。この値を目的変数とする。目的変数に対して、関係ありそうなデータを用意する。（説明変数）

家計調査のデータは、消費支出が500項目あり、各項目の相関を調べやすい。

イ 授業の進め方と学習評価の例

データ例ごとに(ア)～(エ)の授業例とそれに伴う学習評価の例を作成した。

(ア) データの収集

(イ) データの整理（外れ値、異常値）

(ロ) データの特徴（平均値、分散、グラフ等でデータの特徴を捉える）

(エ) データの相関（相関係数、散布図、回帰分析）

ウ ア(ア)のデータを使った授業の進め方と学習評価例を示す

(ア) データの収集

テーマ「2023年の福崎町は本当に暑かったのか？」

2023年8月20日に、福崎町は日本で最高気温となる38.9℃を記録しました。この年の夏のニュースでは、エルニーニョ現象が伝えられました。しかし、2023年の福崎町は本当に暑かったのでしょうか？2022年と2023年の7月と8月の日最高気温（1日における最高気温）の傾向の違いを統計的に分析してみましょう。

学習活動

福崎町の2022年と2023年の7月と8月の日最高気温のデータを気象庁からそれぞれダウンロードし、一つの表（データ）にまとめる。

1. 気象庁から福崎町の2022年と2023年の7月と8月の日最高気温のデータをダウンロードする。

ダウンロード先：<https://www.data.jma.go.jp/risk/obsdl/index.php>

地点を選ぶ：福崎町

項目を選ぶ：データの種類：日別値、項目（気温）：日最高気温

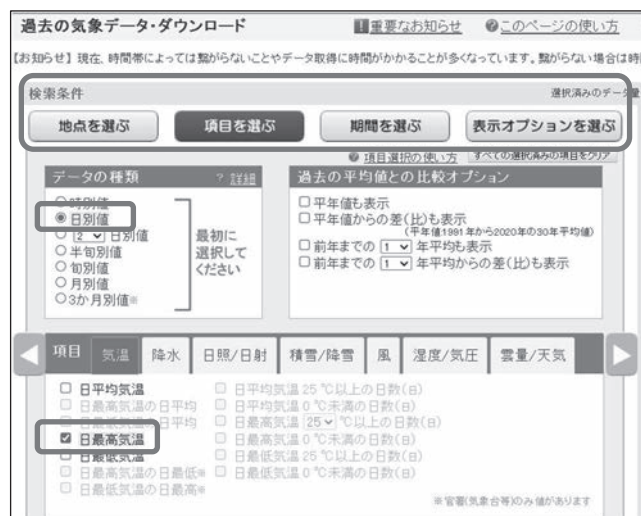
期間を選ぶ：2022年の7月1日から8月31日

2023年の7月1日から8月31日

「過去の気象データ・ダウンロード（地点を選ぶ）」



「過去の気象データ・ダウンロード（項目を選ぶ）」
データの種類「日別値」、気温「日最高気温」を選択



「過去の気象データ・ダウンロード（期間を選ぶ）」

- ・「特定の期間を複数年分、表示する」を選択
- ・「7月1日～8月31日の値を2022年から2022年までを表示」と
「7月1日～8月31日の値を2023年から2023年までを表示」をそれぞれ選択し、ダウンロード

2022年7月と8月の日最高気温のデータをダウンロードしたもの

3				福岡	福岡	福岡	
4	年	月	日	最高気温(°)	最高気温(°)	最高気温(°C)	
5							
6					品質情報	均質番号	
7	2022	7	1	38.2	8	1	
8	2022	7	2	36.1	8	1	
9	2022	7	3	29.4	8	1	
10	2022	7	4	28	8	1	
11	2022	7	5	31	8	1	

2023年7月と8月の日最高気温のデータ

3				福岡	福岡	福岡	
4	年	月	日	最高気温(°)	最高気温(°)	最高気温(°C)	
5							
6					品質情報	均質番号	
7	2023	7	1	25.7	8	1	
8	2023	7	2	32	8	1	
9	2023	7	3	32.1	8	1	
10	2023	7	4	35.1	8	1	
11	2023	7	5	26.1	8	1	

2. 2022年と2023年の7月と8月の日最高気温のデータを一つにまとめる。

(例)

	A	B	C	D
1	月	日	2022年	2023年
2	7	1	38.2	25.7
3	7	2	36.1	32
4	7	3	29.4	32.1
5	7	4	28	35.1
6	7	5	31	26.1

学習評価例

時間	ねらい	重点	備考
1	○問題の解決に必要なデータ 必要なデータの収集について、選択、判断することができる。	思	
2	○尺度の理解 データの尺度水準の違い、「質的データ」、「量的データ」などの扱い方の違いを理解することができる。	知	

重点…重点的に生徒の学習状況を見取る観点を示している。

(イ) データの整理

学習活動

福崎町の 2022 年と 2023 年の 7 月と 8 月の日最高気温のデータを気象庁からダウンロードし、一つの表（データ）にまとめた。このデータの外れ値を求める。

収集したデータをそのまま分析ができるわけではない。データの中には、数値が欠けている欠損値や他の値から大きく外れた外れ値が含まれているかもしれない。分析の行う前に、集めたデータを確認しておくことが重要である。

外れ値

今回は、「第 1 四分位数 $-1.5 \times$ 四分位範囲」より小さい値、「第 3 四分位数 $+1.5 \times$ 四分位範囲」より大きい値を「外れ値」として扱う。

※四分位範囲＝第 3 四分位数－第 1 四分位数

外れ値の求め方

（方法 1）関数から外れ値を求める。

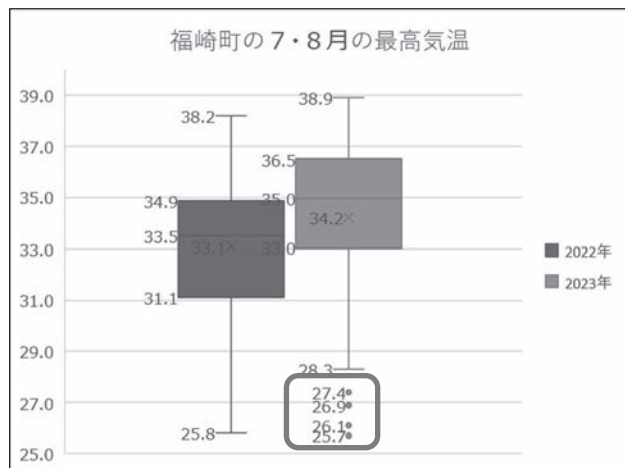
	2022年	2023年
第 3 四分位数 $+1.5 \times$ 四分位範囲	40.5	41.7
第 3 四分位数	34.9	36.5
中央値	33.5	35.0
第 1 四分位数	31.1	33.0
第 1 四分位数 $-1.5 \times$ 四分位範囲	25.4	27.8

中央値（第 2 四分位数）、第 1 四分位数、第 3 四分位数を求めるための数式例

中央値
=median(範囲) または =QUARTILE. INC(配列, 2)
第 1 四分位数
=QUARTILE. INC(配列, 1)
第 3 四分位数
=QUARTILE. INC(配列, 3)

※範囲、配列は、外れ値を求めるための該当データ

(方法2) 箱ひげ図から外れ値を求める。



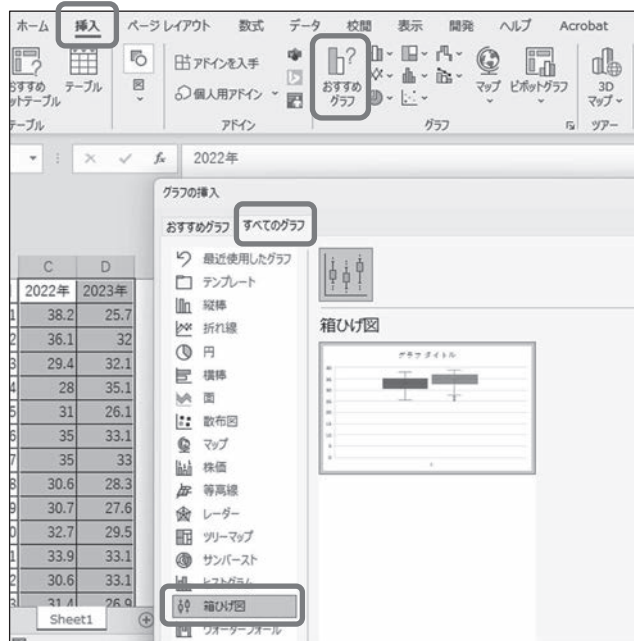
表計算ソフト (Excel 2019) で箱ひげ図を作成する。

1. データを選択する。

B	C	D
日	2022年	2023年
1	38.2	25.7
2	36.1	32
3	29.4	32.1
4	28	35.1
5	31	26.1
6	25	22.1

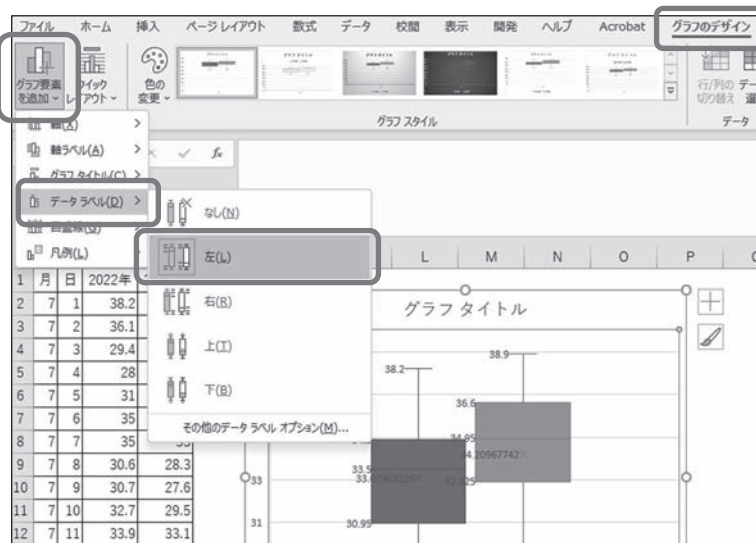
2. グラフを挿入する。

([挿入]-[おすすめグラフ]-[すべてのグラフ]-[箱ひげ図])



3. データラベルを表示する。

([グラフデザイン]-[グラフ要素を追加]-[データラベル])

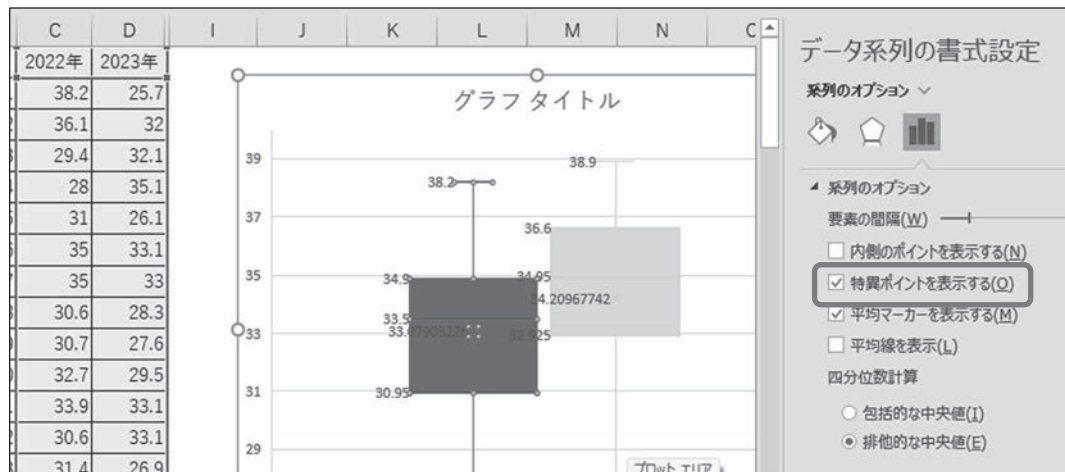


4. 外れ値（特異ポイント）を表示する。

(該当グラフを右クリック-[データ系列の書式設定])



「特異ポイントを表示する」にチェックを入れる。



(留意点)

外れ値を除いて統計処理する場合が多いが、外れ値が入力ミス等による異常値か、ある条件が重なったときに現れる値かは、分析結果を比較した上で判断する必要がある。外れ値を求める他の方法として、クラスター分析などがある。

学習評価例

項目	ねらい	重点	備考
1	○データの整理 データに含まれる外れ値の扱いやデータを整理する必要性を理解し、適切にデータを扱うことができる。	知	

重点…重点的に生徒の学習状況を見取る観点を示している。

(ウ) データの特徴

学習活動

2022年と2023年のデータの特徴を統計量から読み取る。

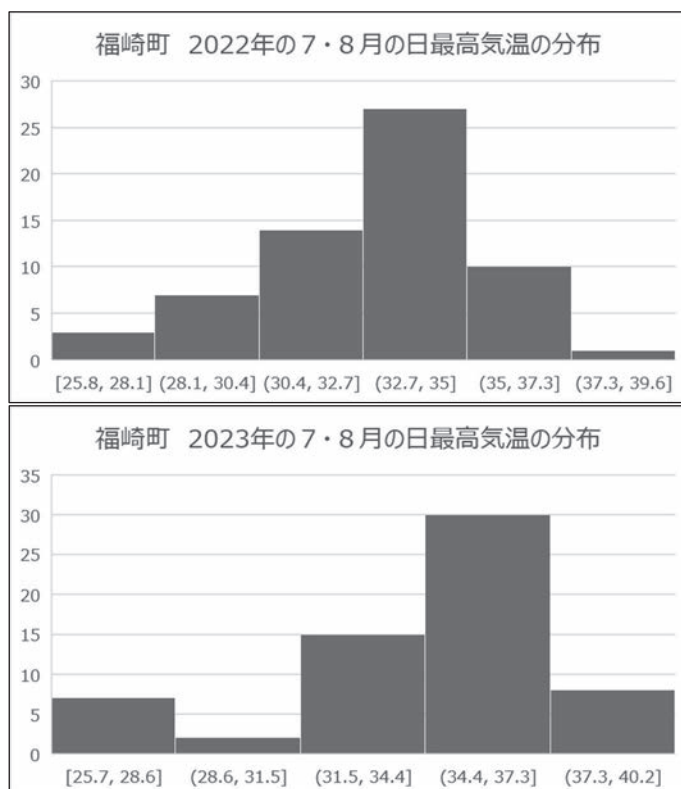
1. 関数から統計量を求め、データの特徴を読み取る。

特徴	統計量	2022年	2023年
分布の状況	最大値	38.2	38.9
	最小値	25.8	25.7
	平均値	33.1	34.2
	中央値	33.5	35.0
散らばり具合	分散	6.4	10.3
	標準偏差	2.5	3.2

統計量を求めるための数式例

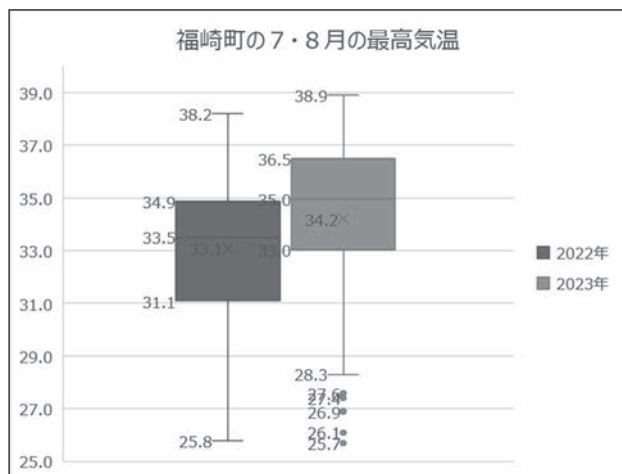
最大値
=MAX(範囲)
最小値
=MIN(範囲)
平均値
=AVERAGE(範囲)
分散
=VAR. P(範囲) (対象にした範囲が母集団のとき)
=VAR. S(範囲) (対象にした範囲が標本であるとき)
標準偏差
=STDEV. P(範囲) (対象にした範囲が母集団のとき)
=STDEV. S(範囲) (対象にした範囲が標本であるとき)

2. 2022 年と 2023 年の 7・8 月の日最高気温の日数を度数分布で表す。



度数分布から、最も頻度が多い（高い）階級（気温）は 2023 年の方が高いことが分かる。

3. 箱ひげ図からデータの特徴を読み取る。



最高気温だけでなく、第1四分位数・中央値・第3四分位数・平均値のいずれも 2023 年の方が高くなっている。このことから、2022 年と 2023 年の福岡市の気温を比較すると、全般的に気温が高い傾向にあったと考えられる。

学習評価例

項目	ねらい	重点	備考
1	○データの特徴を表す指標 ・データの特徴を表す指標とその評価について理解する。 ・データの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。	知 思	

重点…重点的に生徒の学習状況を見取る観点を示している。

(I) データの相関

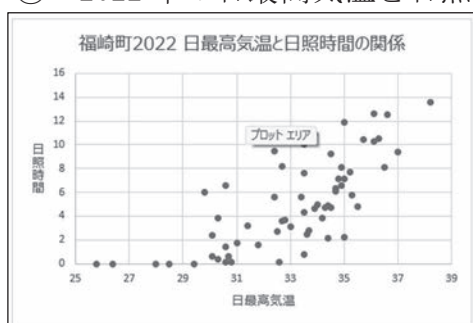
「3 データの特徴」で2023年の福崎町は、2022年と比べ気温が高いことがわかりました。また、気温と日照時間に関係するのかわ確認してみましょう。

学習活動

日最高気温と日照時間の相関を調べる。

1. 日最高気温と日照時間の相関を調べる。

① 2022年の日最高気温と日照時間を散布図で表す。



② 相関係数 (Excel 分析ツールの利用) を求める。



相関係数の結果

最高気温(°C)日照時間(時間)	
最高気温(°C)	1
日照時間(h)	0.722548

相関係数の結果、日最高気温と日照時間が互いに関係があることがわかる。

相関係数の判断基準（効果量）

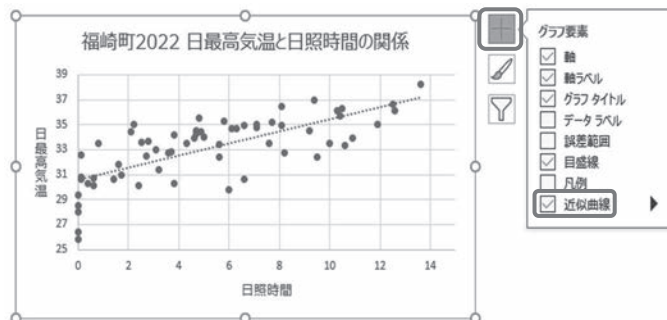
効果量例

相関係数の値	相関係数の強弱
1～0.7	強い正の相関
0.7～0.4	中程度の正の相関
0.4～0.2	弱い正の相関
0.2～－0.2	相関がない
－0.2～－0.4	弱い負の相関
－0.4～－0.7	中程度の負の相関
－0.7～－1	強い負の相関

相関係数は、各項目に何かしら関係があるかを調べるもの
必ずしも因果関係（要素同士が原因と結果の関係）を表すものではない。

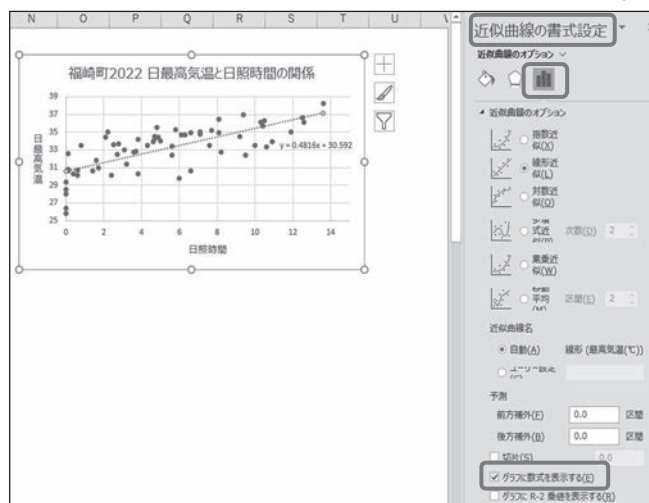
2. 予測を立てる。（日照時間によって、日最高気温を予測する）

① 近似曲線の追加



② 近似曲線の数式を表示する。

近似曲線の書式設定から数式を表示させる。



$y = 0.4816x + 30.592$ という予測式が成り立ち、 x (日照時間)により、 y (日最高気温)の予測した値が求められる。

学習評価例

項目	ねらい	重点	備考
1	○相関（相関係数、散布図） - 一連のデータの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する。 - 客観的な指標を基にデータの傾向を判断し、自身の考えを基に適切な解釈を行うことができる。	知 思	
2	○相関の有無の判断 - 客観的な指標を基に、相関の有無を判断することができる。 - 客観的な指標を基に、データの傾向を評価し、適切な解釈を行おうとする。	思 態	
3	○相関と因果 データの特徴を表す指標と、その評価について理解することができる。	知	
4	○仮説とその検証に必要なデータの考察 データを問題の発見・解決に活用するために、必要なデータの取扱いを理解する。	知	

重点…重点的に生徒の学習状況を見取る観点を示している。

おわりに

本研究の目的は、高等学校で行うデータ活用分野の授業を通して育む力を整理するとともに、演習で使用するデータを作成するための視点を提案することである。当サイトを活用し、生徒にどのような力（データの特徴を捉える、データの相関を見つける等）を身に付けさせたいのかを考え、授業で用いる教材のデータを工夫することが、生徒一人一人にデータの分析能力の育成することにつながる。

今後も、作成した当サイトについて、研修講座等を通じて、より広く周知するとともに、学校からの意見を基に改善を図り、充実させていきたい。

謝辞

本研究に御協力いただいた皆様に心より感謝を申し上げます。