

探究活動を支えるデータサイエンス入門

問いを立て、データで確かめ、次の問いにつなげる

Web用

日出学園中学校・高等学校
武善 紀之

<https://www.hinode.ed.jp/share/takeyoshi/>



自己紹介

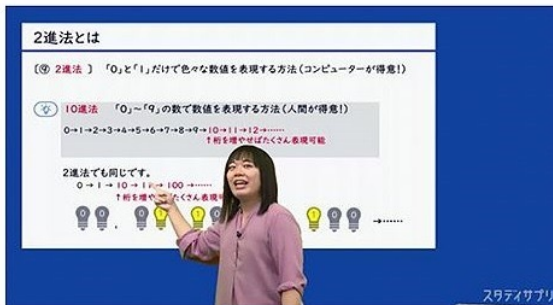
名前 武善紀之(たけよし のりゆき)

所属 私立 日出学園中学校・高等学校 教諭

所在地:千葉県市川市 ※幼稚園・小学校併設

教科 情報科 13年目(+数学,公民,技術,社会,保健)

スタディサプリ



NHK
高校講座



武善 紀之
たけよし のりゆき
日出学園中学校・高等学校教諭



<https://www.hinode.ed.jp/share/takeyoshi/>

本日の内容(事前告知)

- 授業や探究活動を通じた生徒の情報活用能力の育成に向けて、教員のICTを活用した指導力の向上に資する。
- 探究活動において必要となる、データの整理・可視化・比較・分析の基本を、ICTを活用して実践するための講義
- Google フォーム・Google スプレッドシートを用いたアンケート調査の設計・実施・分析。設問設計上の注意点、代表値の算出、相関係数の確認、グラフ化を扱うとともに、相関と因果の違いや仮説検定の基本的な考え方にも触れる。
- 情報Ⅰ・情報Ⅱの内容を踏まえ、単回帰分析、クラスタリング、主成分分析など、より発展的な分析手法を、オープンデータも活用しながら体験的に学ぶ

今日、この研修で目指すこと

- ×難しい統計手法を教える研修
 - ×自分で主成分分析ができる先生になる研修
 - 探究をデータで支えるための地図を頭に入れる研修
(正直、今は生成AIが分析はかなりやってくれます)
-

生徒の探究に「次の一手」を返せるようになる

そのために今日は……

- ・自分で一度、問い → データ収集 → 分析 → 結論まで体験する
- ・データを扱うときの落とし穴や限界を知る
- ・「この問いなら、こんなデータ・分析が使えるそう」と考えられるようになる

Agenda.

1. 「探究学習」と「統計」の関係
2. データサイエンスの基本(PPDACサイクル)
3. 実際の指導例(ポスター作成、プレゼン、相関分析)
4. 研究上の落とし穴(アンケート調査、相関と因果)
5. ステップアップした分析(仮説検定、多変量解析)

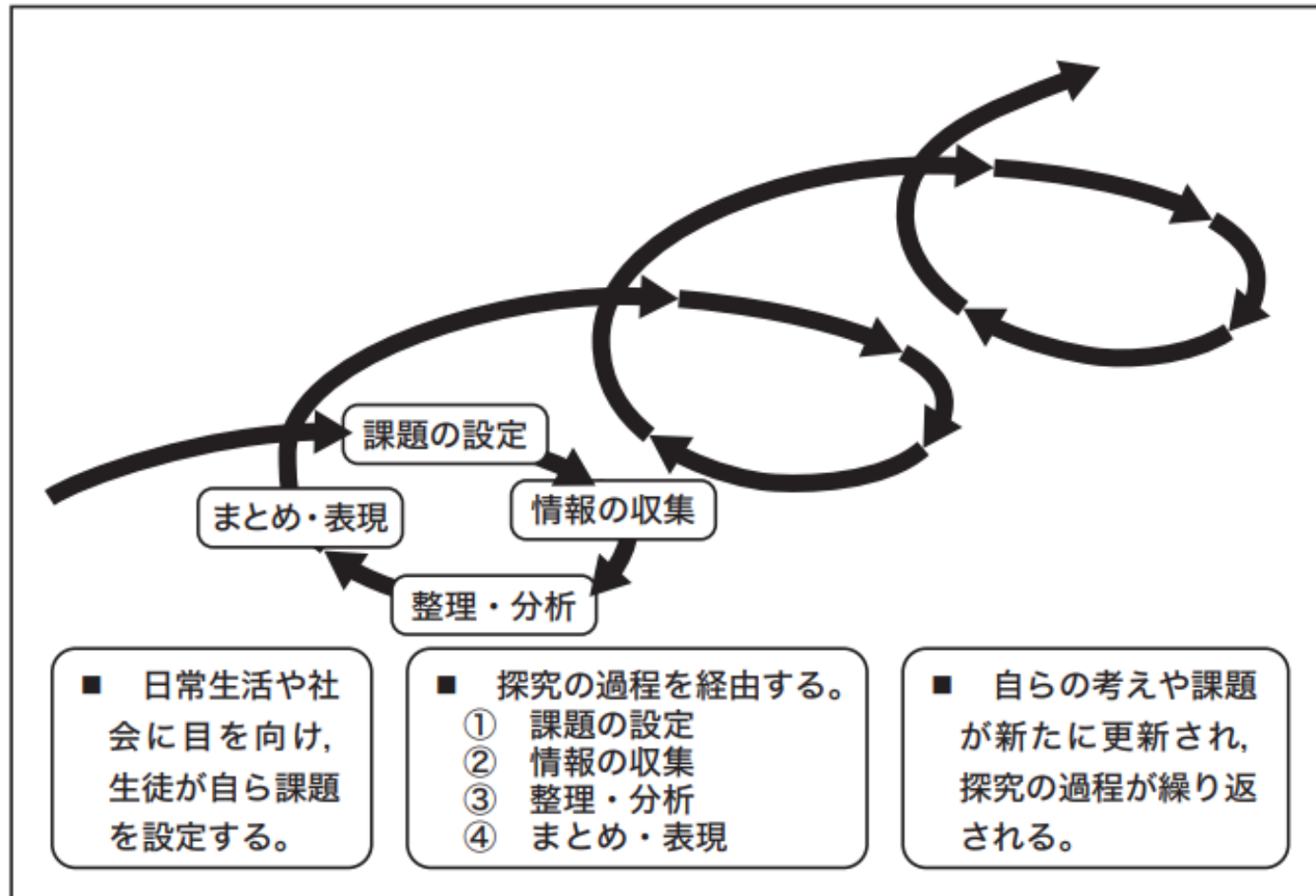
進行イメージ案

時間	内容
10:05～10:20	1.「探究学習」と「統計」の関係
10:20～10:50	2. データサイエンスの基本 アンケート作成→休憩10分(アンケート回答)
11:00～12:00	2. データサイエンスの基本 データの分析以降 3.実際の指導例
12:00～13:00	昼休み
13:00～13:30	4. 研究上の落とし穴 5. ステップアップした分析 ①数学 I とつなげよう
13:30～14:40	5. ステップアップした分析 ②以降
14:40～15:00	まとめ・質疑

1. 「探究学習」と「統計」の関係

探究学習の形

探究における生徒の学習の姿



①【課題の設定】

体験活動などを通して、課題を設定し課題意識をもつ

②【情報の収集】

必要な情報を取り出したり収集したりする

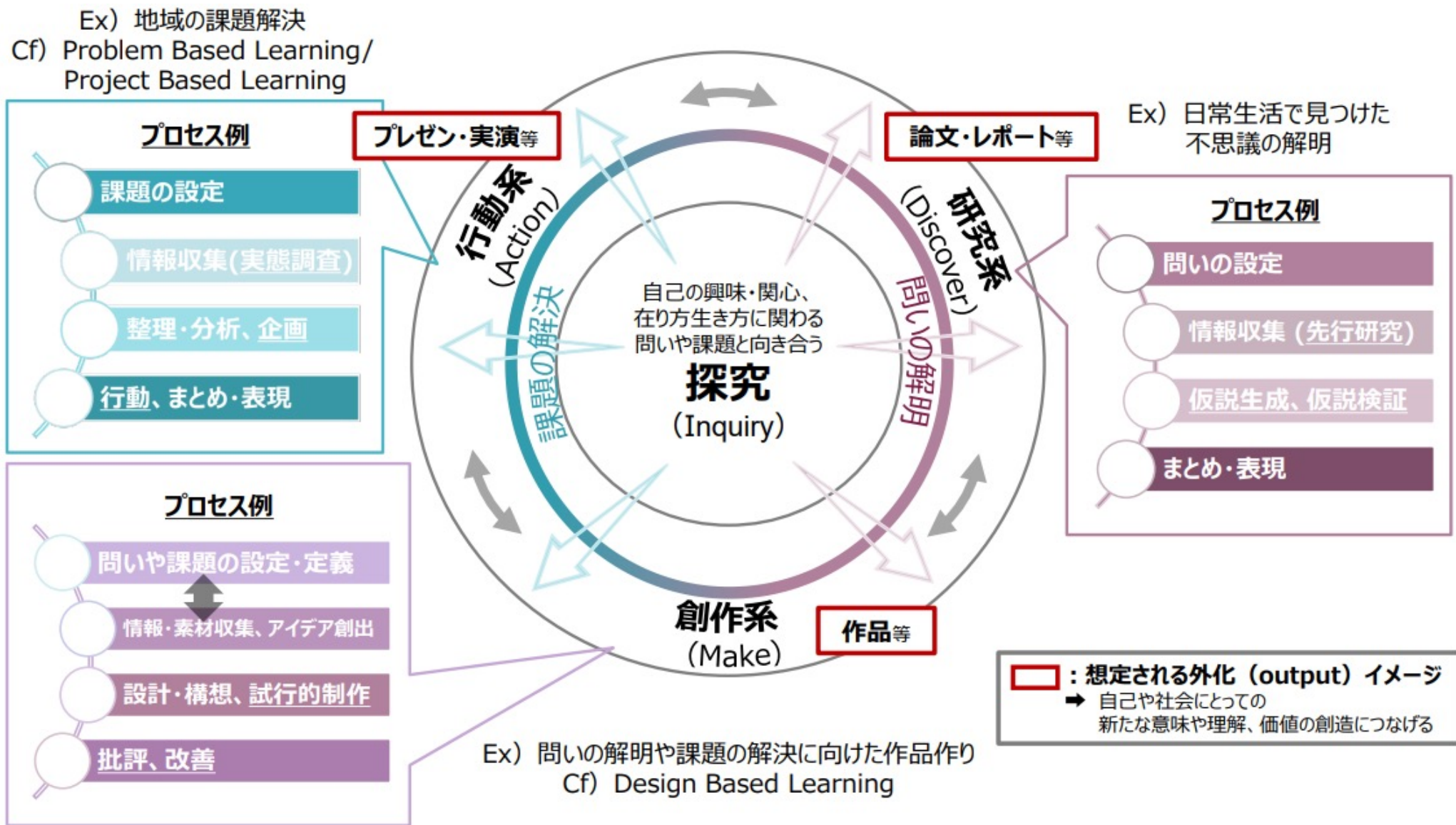
③【整理・分析】

収集した情報を、整理したり分析したりして思考する

④【まとめ・表現】

気づきや発見, 自分の考えなどをまとめ, 判断し, 表現する

次期学習指導要領での「探究」



わかりやすく表現すれば……



東京書籍「新編情報Ⅰ」紙面より

実は「情報科」と「データサイエンス」は関わりが深い

情報科

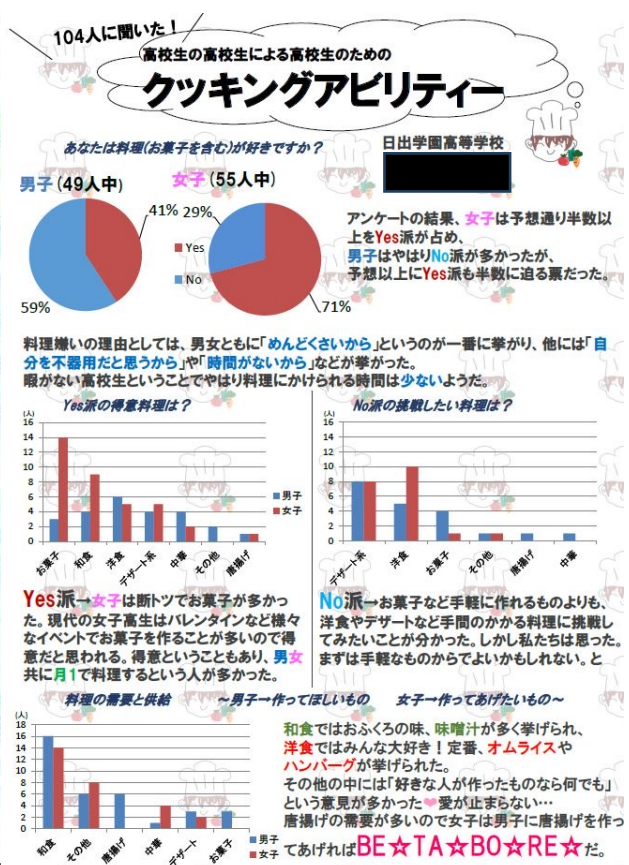
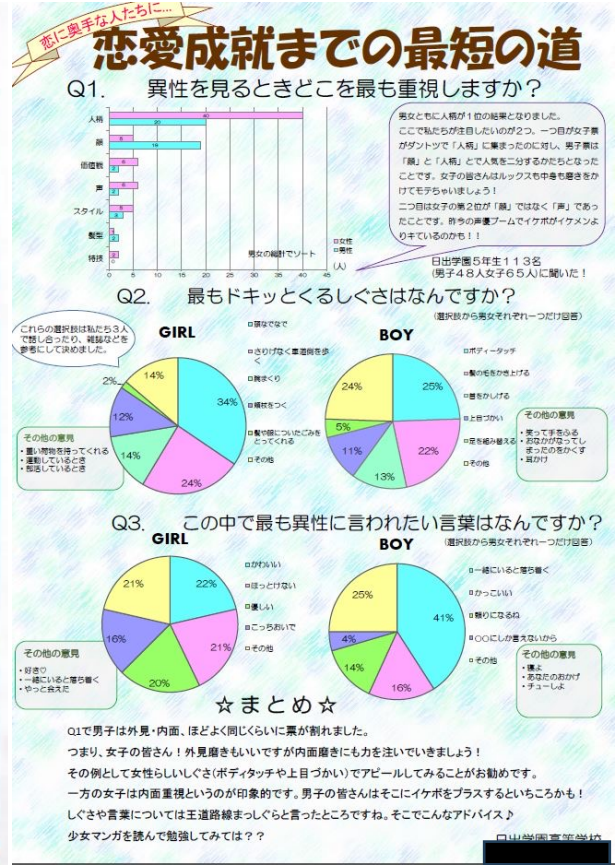
情報Ⅰ 「情報通信ネットワークとデータの活用」

情報Ⅱ 「情報とデータサイエンス」

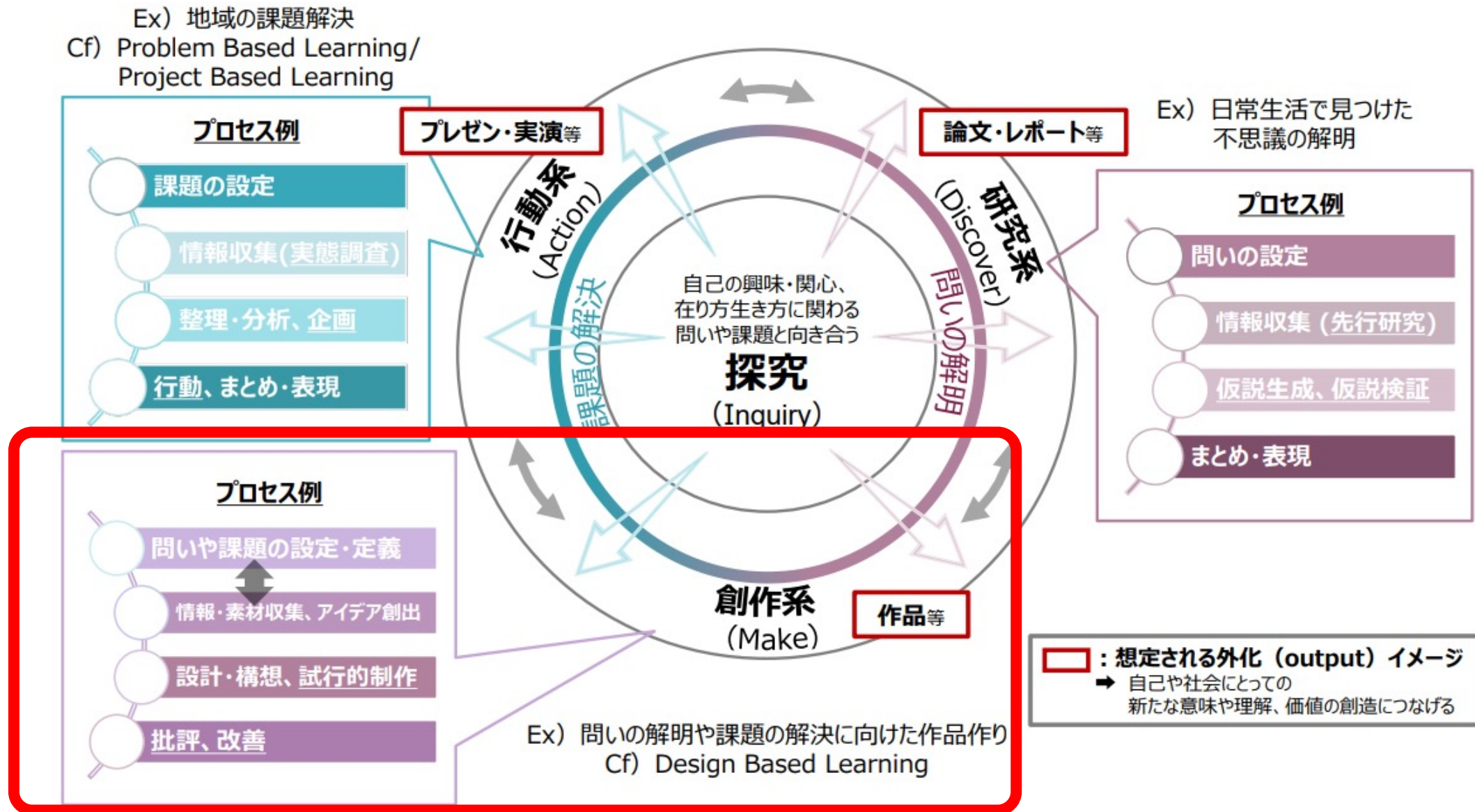
文科省学習動画 <https://nttls-edu.jp//joho/>



～2019年まで「情報科」で実施（以後、「総合的な探究の時間の実習」へ統合。現在は未実施）



「研究系」以外でも「統計」は活用される



実は「創作系」でも「データ分析」は絡んでくる



- 情報Ⅱの授業で「身の回りの問題」を解決する「アプリ」を作成。
- アプリの改善に向けて「ユーザー調査」を実施。
- アンケート結果に基づいて、アプリを改善。

データで考えるのは、理系だけではない

人や社会を扱う学問でも、問いに答えるためにデータが使われている

経済学

物価・賃金・雇用などから、社会の動きや政策の効果を捉える。

計量社会学

調査や統計から、格差・家族・教育など、見えにくい社会の構造を浮かび上がらせる。

文学研究

語の頻度や組み合わせから、作家や時代による表現の特徴を読み解く。

歴史学

古文書・新聞・人口記録などから、社会の長期的な変化を捉える。

言語学

大量の文章や会話を集めたコーパスから、言葉の実際の使われ方を調べる。

政治学

選挙結果や世論調査から、人々の意識と政策との関係を分析する。

心理学

実験や調査を通して、人の感情・判断・行動の傾向を確かめる。

法学

判例や裁判結果を分析し、法律が社会でどう運用されているかを検証する。

「文系の問い」に、「データという方法」で迫る。

「データ活用」は「数学」「理系」だけの話じゃない

現状を捉え、問いを立て、試し、確かめ、より良くしていくための方法

生徒の立場

「変えたい」を、
根拠ある提案へ

- 校則を変えたい → 誰が、何に困っている？
- 行事を良くしたい → 参加状況や声から課題を探る
- 主張を伝えたい → 「そう思う」から根拠を示すへ



教員の立場

「うまくいかない」を、
検証できる問いへ

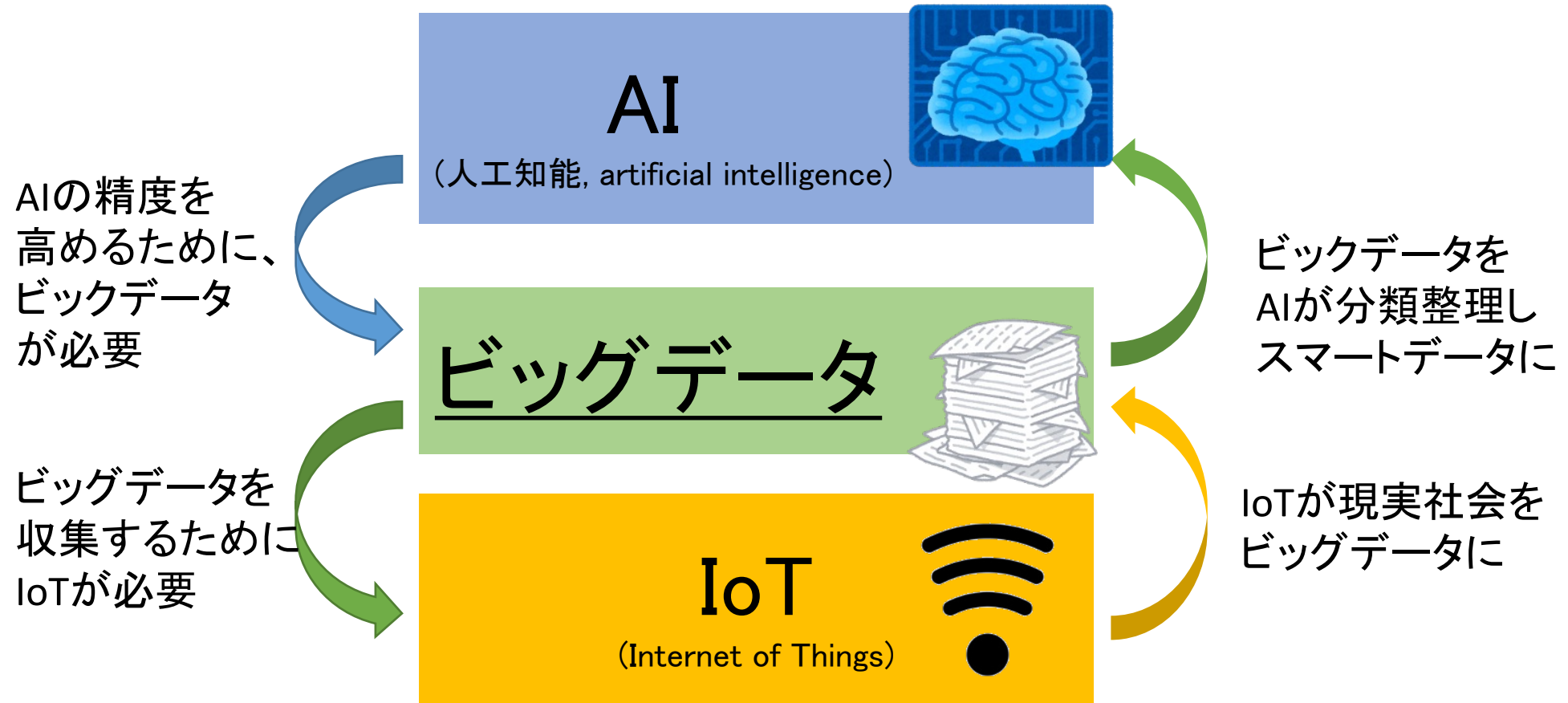
- 成績が伸びない → どこでつまづいている？
- 新しい授業方法 → 実施前後で何が変わった？
- 経験や勘 → 仮説としてデータで確かめる



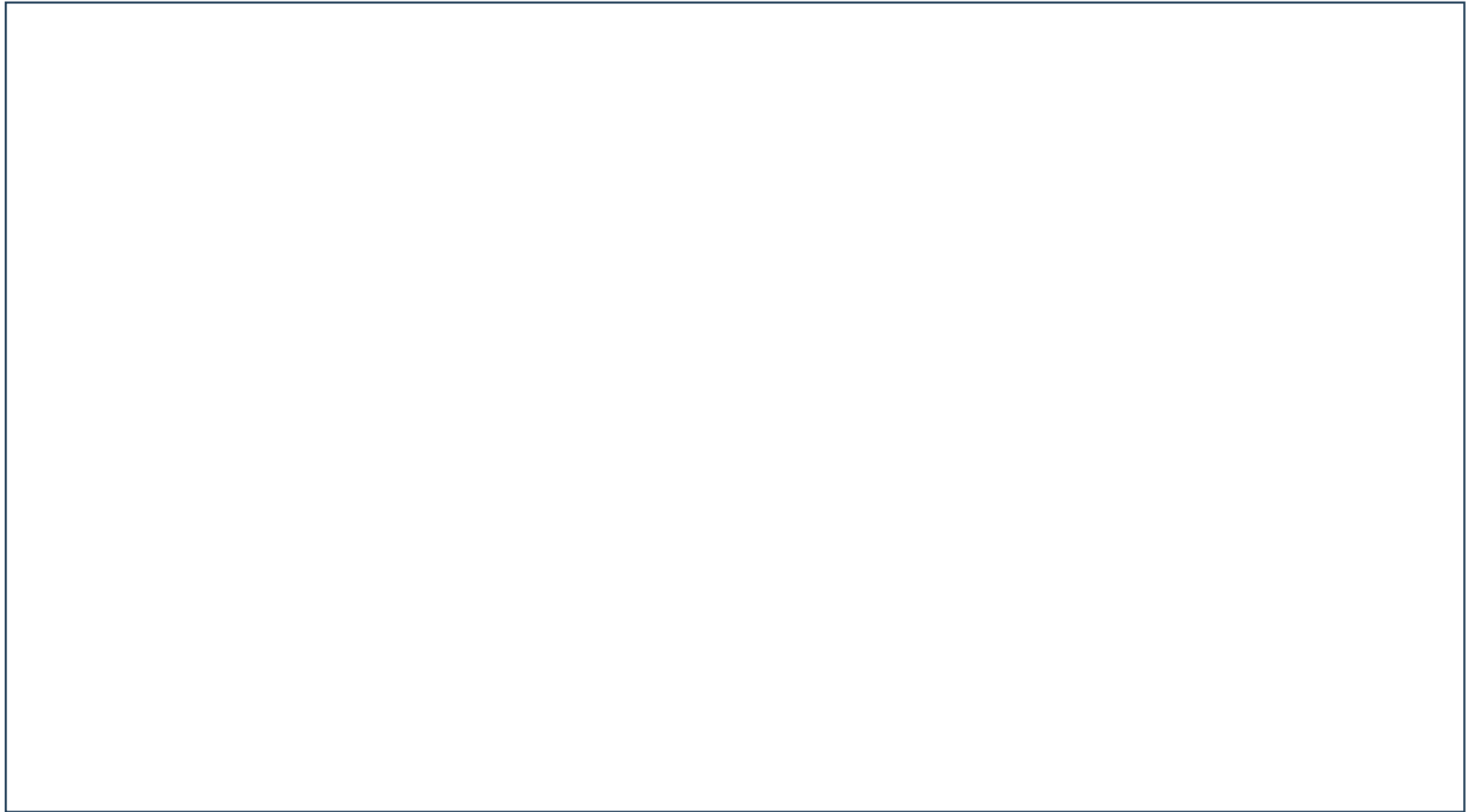
数学や統計は「道具」。出発点は「知りたい・変えたい・確かめたい」という思い。

社会の仕組み(データ駆動型社会)

データは21世紀の**石油**である。



データサイエンス系学部の新設も相次ぐ



朝日新聞Thinkキャンパス「新設ラッシュ『データサイエンス学部』なにを学ぶの？ 卒業後の進路は【親のための学部・学科講座】」、2023年4月20日、閲覧日：2026年9月21日、

<https://www.asahi.com/thinkcampus/article-100311/>

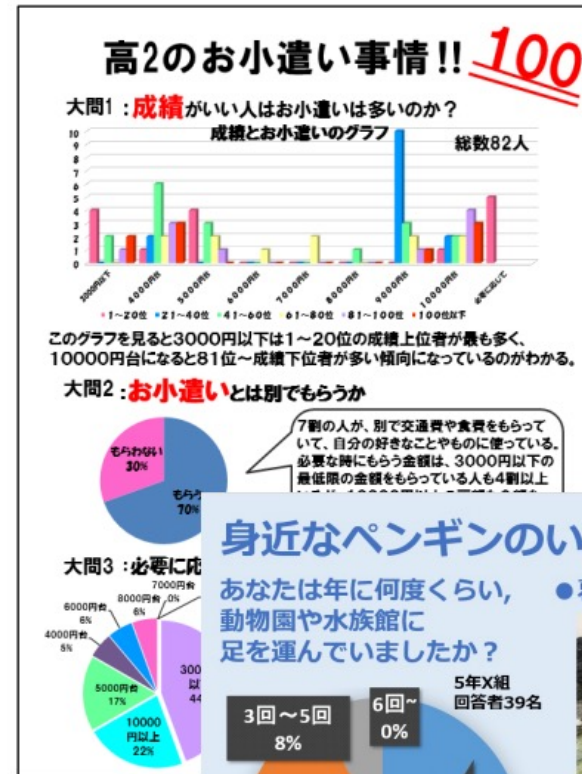
結局のところ、データサイエンスとは？

「身近な疑問」「思い」の「見える化」

「身近な疑問」や
「思い」を



「見える化」して解決！



身近なペンギンのいる水族館(都内)

あなたは年に何度くらい、
動物園や水族館に
足を運んでいましたか?

●葛西臨海水族館

●江戸川区動植物園



●サンシャイン水族館

●すみだ水族館



なんと無料!

今日の内容 「生徒の立場」で一週 + 指導のコツ

① 問いを立てる

「何を知りたい？」を考える



② データを集める

実際にアンケートを作って回答する



③ データを分析する

グラフ・平均・相関などを実際に試す



④ 結論を考える

結果から何が言えるか、次に何を調べるか



⑤ もう一段深く見る

バイアス・因果・仮説検定・回帰など

午前

(10:25～12:00)

午後(13:00～15:00)

2. データサイエンスの基本

※以下の動画に準拠します(事後の活用にも向けて)



<https://www.youtube.com/watch?v=VFzCUPp30jg&t=22s>

https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_01832.html

文科省Webページにスライド、ワークシート等もあり。

PPDAC サイクルを意識したアンケート実習	
データの活用②	
組 番 氏名	
1. ねらい PPDAC サイクルが利用して、統計的な手法を使った問題解決をしよう。 [P] () 問題を明確にして、仮説を立てる。 [P] () 仮説を検証するためにはどのようなデータを集めればよいかを考える。 [D] () データを収集し、必要に応じて表などに整理する。 [A] () 仮説を検証するためにグラフなどを作成し、データの傾向を把握する。 [C] () 仮説が正しかったか考察する。	(6) アンケート結果
2. アンケート調査をしよう。下記に当てはまる項目を書きましょう。 (1) アンケートの目的 (2) 仮説 (3) 仮説の理由 	(7) 考察
(4) アンケートの質問文 	(8) 結論
(5) アンケート結果の分析方法 	3. 次の PPDAC サイクルに向けての今後の課題を書きましょう。

データの活用②

PPDACサイクルを意識した
アンケート実習

何事にも“型”がある。

cf.PDCAサイクル



PDCA サイクル：
問題解決に用いられるプロセスの1つ。
Plan（計画），Do（実行），Check
（評価），Act（改善）を何度も繰り返
し回していく。

大事なことは,ぐるぐると回すこと!
問題解決が終わったら,次の問題解決へ!

統計調査のPPDACサイクル

多くの生徒が、勉強のせいで睡眠不足に陥っているのかも！
今後、改善したい！

結論
Conclusion

C

学習時間が多い人は、睡眠時間が短い！

分析
Analysis

A

問題の決定
Problem

P

睡眠時間と
学習時間に関係性があるか？

計画
Plan

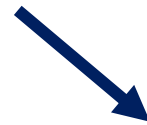
P

クラスに、
アンケート調査を
やってみよう

データ収集
Data

D

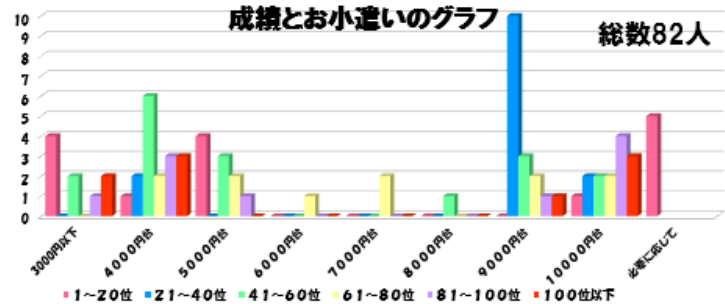
アンケートを実施



先程のポスターもPPDACサイクルを意識

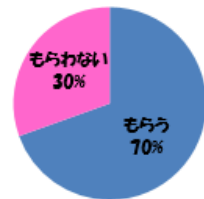
高2のお小遣い事情!! 100

大問1: **成績**がいい人はお小遣いは多いのか?



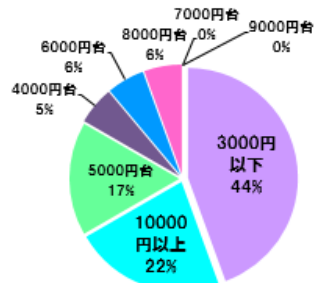
このグラフを見ると3000円以下は1~20位の成績上位者が最も多く、10000円台になると81位~成績下位者が多い傾向になっているのがわかる。

大問2: **お小遣い**とは別でもらうか



7割の人が、別で交通費や食費をもらっていて、自分の好きなことやものに使っている。必要な時にもらう金額は、3000円以下の最低限の金額をもらっている人も4割以上いるが、10000円以上の高額な金額をもらっている人も2割以上いるという対照的な結果になった。

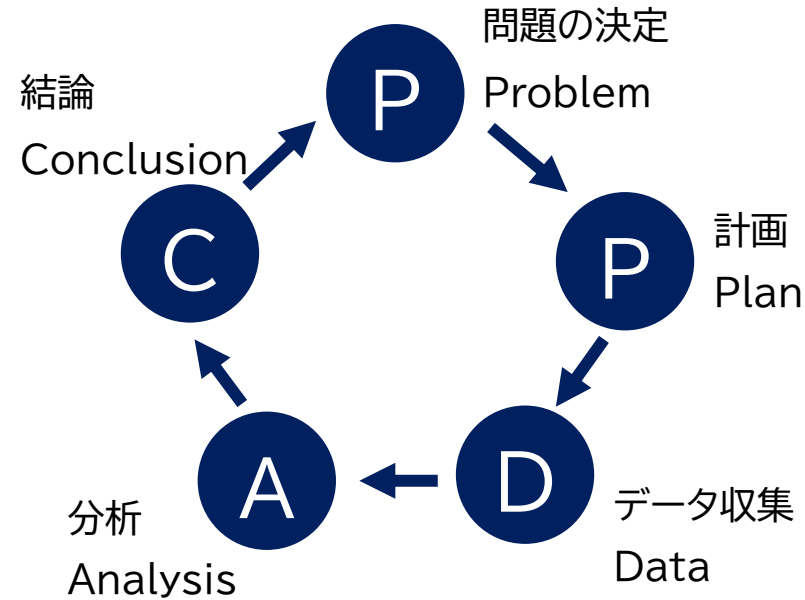
大問3: 必要に応じてもらう人の**金額**



5年2組

Point

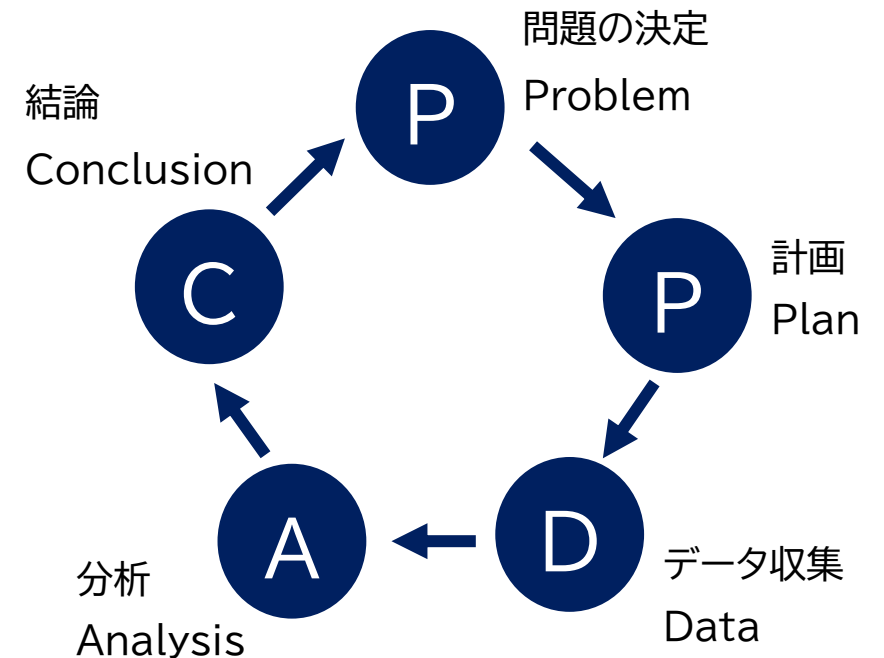
予想では、成績上位者ほどもらっていると思ったが、予想と反する結果となった。親が甘やかしているという事が成績低下へと導いているのかもしれない。。。



- **問題** (もっとお小遣いが欲しい!)
- **計画** (成績と関係性?)
- **調査** (クラスアンケート)
- **分析** (むしろ下がる!)
- **考察** (甘やかし・・・?)

(こぼれ話)

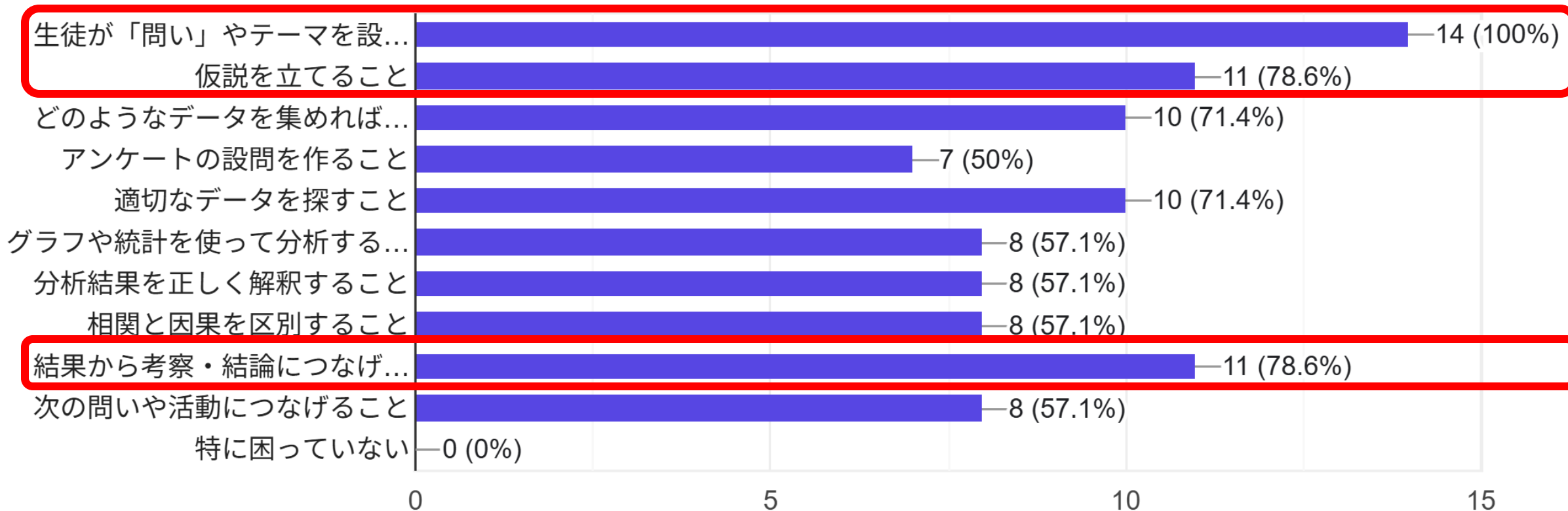
- 統計と言われると…「A(分析)」にばかり注目が集まりがち
- しかし実際は、生成 AI も登場したことにより、A(分析)はより簡単になりつつある(実例で示すと……)。
- 指導の上で、実は大変なのが他の部分、特にProblem(問題)、Plan(計画)、結論(Conclusion)。



事前アンケートの結果より

10. 生徒の探究活動を指導する際、難しい・困ると感じることをすべて選んでください。

14 件の回答



- 生成AI時代、要求されることは生徒のファシリテート(伴走)。
- そのためにも、A(分析)にどんなものがあるかの見通しを持っていることは結局必要(特に教員)。

D(Data)には大きく2通りのやり方がある

① アンケート調査実習→「自分の問いに合わせて集める」

- 自分が知りたいことに合わせて質問を作れる
- 生徒自身が「問い→収集→分析」の流れを体験しやすい
- 学校・学年・クラスなど、身近な対象を調べやすい
- ×回答数が少ないと、一般化しにくい
- ×作成・回答に時間を取られる

② オープンデータの分析実習→「すでにあるデータから問いに迫る」

- 公的機関などのデータなら、比較的信頼性が高い
- ×自分の問いにぴったり合うデータがあるとは限らない
- ×データの加工・処理が難しいことがある。

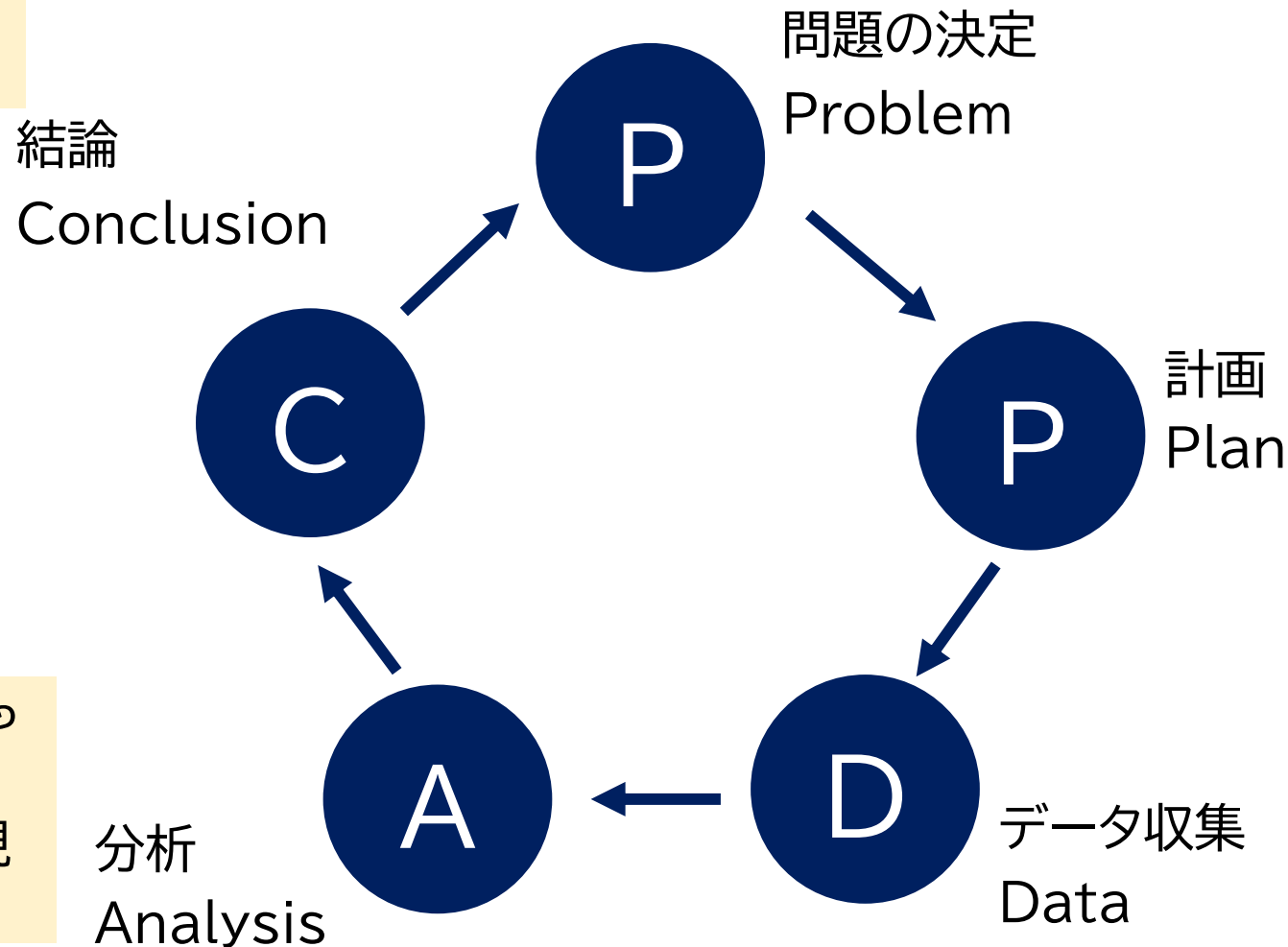
オープンデータとは？

- 国や地方公共団体、民間企業などが保有する公共データを、誰もが無償で機械判読に適した形式で、営利・非営利を問わず二次利用できるルールのもとでインターネット等に公開したもの
 - [青い森オープンデータカタログ](#) : 青森県のオープンデータカタログ
 - [e-Stat 政府統計の総合窓口](#) : 政府のオープンデータ集(難しい)
 - [統計局ホームページ/令和3年社会生活基本調査](#)
 - [SSDSE\(教育用標準データセット\)](#) : e-Statの加工済データ
 - [Let's Start!!! ~e-Stat活用ナビ~](#) : 活用事例集
 - [統計ダッシュボード](#)
 - [RESAS 地域経済分析システム](#)
- 最近は大いび使いやすくなった
- オープンデータを2次的にまとめたサイトも使える。
 - [都道府県別統計とランキングで見る県民性 - とどらん](#)

オープンデータを活用する場合のPPDAC

人口減少の進み方は地域によって違う。対策も一律ではなく、地域ごとに考える必要がある

青森県では、人口減少や高齢化はどの地域で特に進んでいるのか？



e-Statや自治体のオープンデータを使って、市町村ごとの違いを調べてみよう

地域ごとの増減率や年齢構成を比較し、地図やグラフで可視化する

人口、年齢構成、転入・転出、出生数などの公開データを集める

今回は「①アンケート調査実習」を試します

① アンケート調査実習→「自分の問いに合わせて集める」

- 自分が知りたいことに合わせて質問を作れる

- 生徒自身が「問い→収集→分析」の流れを体験しやすい

- 学校・学年・クラスなど、身近な対象を調べやすい

- ×回答数が少ないと、一般化しにくい

- ×作成・回答に時間を取られる

② オープンデータの分析実習→「すでにあるデータから問いに迫る」

- 公的機関などのデータなら、比較的信頼性が高い

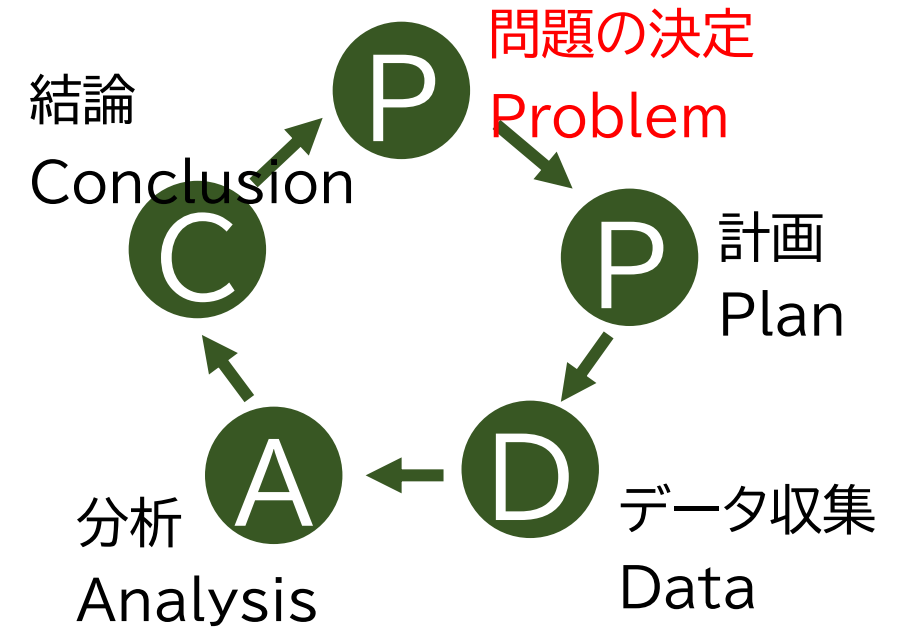
- ×自分の問いにぴったり合うデータがあるとは限らない

- ×データの加工・処理が難しいことがある。

PPDACサイクルに沿って、
アンケート実習を行ってみましょう。

クラスの実態調査をやってみよう！





Problem 問題の発見

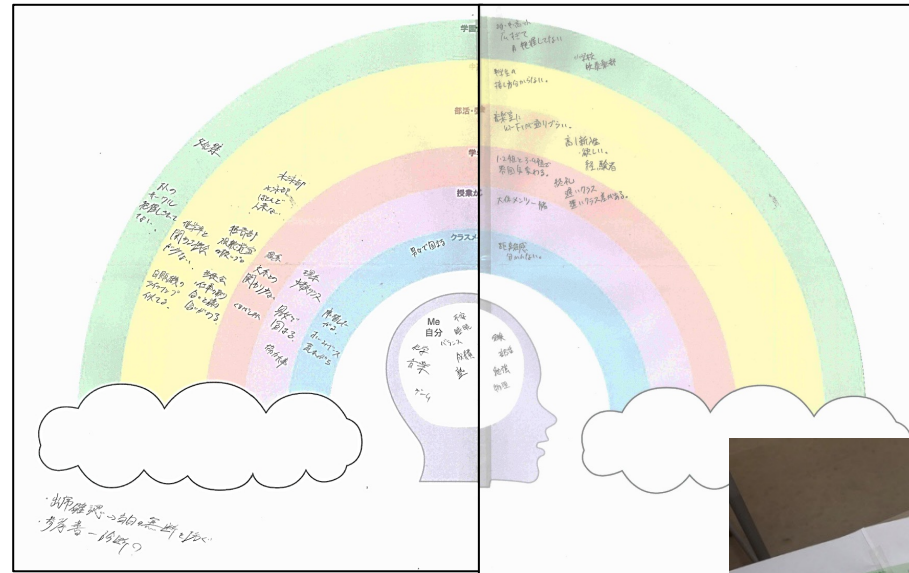
Problemの見つけ方

- 情報科でもここが一番の難所。「問題解決」<「問題発見」
- 最近「アプリ開発」の授業でやったこと(Be you sheet)



Swift Students Community Japan

<https://swiftstudents.jp/about/beyousheet>



Problemの見つけ方

- 生徒の実態によっては「仮説発見型」のPPDACでも良いのかもしれない。
- 「まずやって、やった中で仮説を見つけていく」という作り。

仮説検証型の探究

朝食を毎日食べる生徒の方が、学力が高いのではないか？
(メリット: PPDACが明確・ゴールがハッキリする)

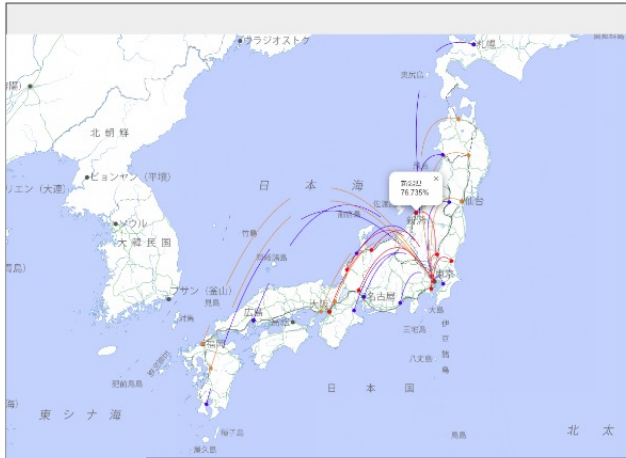
仮説発見型の探究

学力が高い生徒には、生活習慣上どんな特徴があるのだろうか？

- 色々調べてみて、「睡眠時間」と関係がありそう、を見つけていく。
- では、なぜ「睡眠時間」と関係があるのか？を深掘り。

Problemの見つけ方

- オープンデータに、先に色々触れてみても「やってみたい」テーマが見えてくるかも。



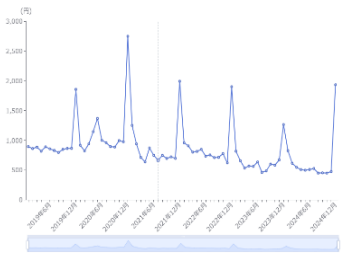
調べたデータ
（「どこ」の「何」？URLと一緒に。
他の気になった人も調べられるように！）

RESA 日本の米菓の生産割合
<https://resas.go.jp/consumption-pos-from-to/?pref=13&tab=0&year=2024&month=12&level=pref&lat=35.96911507577482&lng=138.60351562500006&mapTile=pale&indicator=to&broad=10&middle=130100&small=130129>

判明したこと・その事実から考えられること

生産地の割合はコメの生産量の多い新潟県がちぎりで1位だった。東北地方に集中している

調べてみたこと レジを通過した1000人あたりの冷凍食品の購入金額



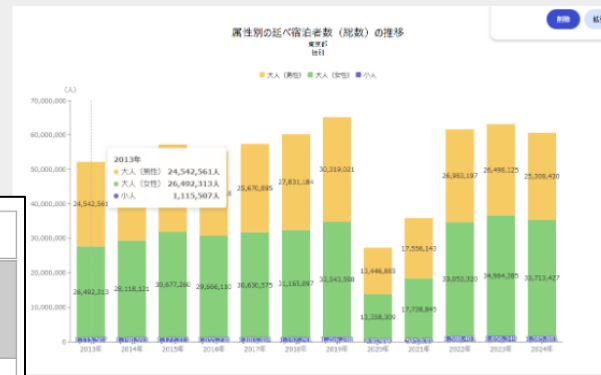
調べたデータ
（「どこ」の「何」？URLと一緒に。
他の気になった人も調べられるように！）

RESASの生活用品消費分布の項目から
https://resas.go.jp/consumption-pos-trend/?tab=1&level=pref&pref=13&year=2024&month=12&broad=110000&middle=110700&small=110701&categoryLevel=small&transitionContent=pi_price&barContent=consumption&rankingMethod=pi_ratio

判明したこと・その事実から考えられること

- 毎年12月になると2, 3倍に跳ね上がる
→お正月に向けてカニやエビなど高価なものを購入する傾向があるため
- 2020年はコロナの影響で自宅で食事をする人が多くなり需要が高まったため高くなっていると考えられる。

調べてみたこと 東京都の宿泊者数の推移



調べたデータ
（「どこ」の「何」？URLと一緒に。
他の気になった人も調べられるように！）

<https://resas.go.jp/tourism-guest/?pref=13&city=13101&tab=3&year=2024&level=pref&lat=37.58811876638322&lng=142.73437500000003&zoom=6&mapTile=pale&opacity=0.8&sort=0&displayMethod=1&displayCategory=11>

判明したこと・その事実から考えられること

- 新型コロナの影響で2020、2021年は減っている。
- 毎年、男性より女性の人の人が多いのは、

日出学園中学校・高等学校「【情報科】『こんな関係性がありそう！』データサイエンス入門」日出学園、2026年2月22日、閲覧日：2026年9月21日、
<https://www.hinode.ed.jp/high/blog-h/36657/>

なにを調べるか？

- アンケートの目的
 - 高校生の実態を明らかにしよう
- 仮説
 - 高校生の実態の予想
- 仮説の理由
 - 「仮説」のようになると考える理由

統計データ等の根拠も
探してみよう

本日はせっかくなので、
「青森県の教員の实態」というテーマで、
集まっている先生方を対象に、
実態調査をしていただこうと思います！

Try I . 何を調べるか？

1. ○○高校 ○○○○先生

仮説	○○○○
仮説の理由	ああ
聞くべきこと	ああ

結果
(スクショで挿入)

結果・考察

次のサイクルで扱いたいこと

どんなテーマでアンケート実習をするか、
決めてみましょう！

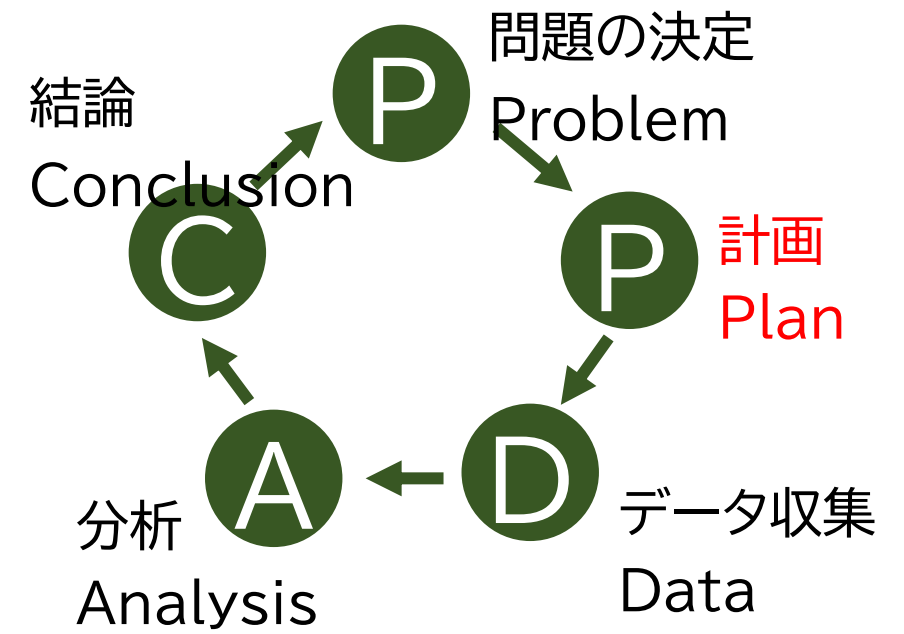
(例)

学習時間と睡眠時間の間に、関係性はあるの
だろうか？

→ (仮説) 学習時間が長いほど、睡眠時間は短い。

→ (理由) 勉強をよくする人は、睡眠時間を削って勉強していて、

寝不足な人が多いと考えたから。



Plan 計画

どのようなデータを集めればよいか？

予想するアンケート結果を得るために・・・

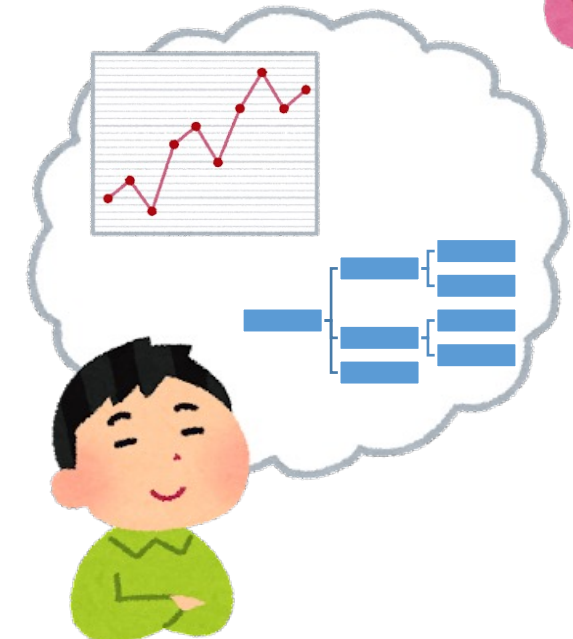
- **どんなデータが欲しい？**

- 量的データ：数量の大きさを表す
- 質的データ：種類の違いや区別を表す

- **どんな分析をしたい？**

- 定量分析：ものごとの状態を、数量を使って分析する（５段階評価など）
- 定性分析：ものごとの状態を、数量以外で分析する（自由記述など）

- **どのように可視化する？**



どのようにデータを集めればよいか？

アンケート文の設計は大切！

Q. 何時間勉強していますか？

- ① 0～1時間
- ② 1～2時間
- ③ 2～3時間
- ④ 3時間以上

学校の授業は？

1時間の人は、
①と②のどちら？

平日？休日？

週あたり？1日？



Q. あなたは平日1日に、平均で何時間勉強をしていますか？（学校の授業以外、塾等も含む）

- ① 1時間未満
- ② 1時間以上2時間未満
- ③ 2時間以上3時間未満
- ④ 3時間以上

一番良いのは、「隣の人」に試しに答えてもらうこと！（自分では気付けない）

アンケート調査は意外と難しい

統計調査ガイダンス ②アンケートの作成

【アンケートの設計】

アンケート調査は意外に、難しい！ 次の設問の問題点を考えてみよう。

Q. 何時間勉強していますか？

1 : 0～1時間 2 : 1～2時間 3 : 2～3時間 4 : 3時間以上

【答えにくいアンケートにしないコツ】

(1) ことばづかい・文章表現(wording)

ex. ヒゲペンギンとマカロニペンギンのどちらが好きですか？（みんなはペンギンに詳しくない）

ex. 「宿題は少ないが、いつも冷たい先生」と「いつも冷たいが、宿題は少ない先生」

(2) あいまいな表現

ex. ペンギンはよい生きものであると思いますか？（よい生き物って・・・？）

(3) ダブル・バーレル質問（二重質問）(double-barreled question)

ex. 「あなたは日本食や日本のアニメが好きですか？」（日本食好きだけどアニメ嫌いな人は？）

【アンケートの種類】

(1) 単一選択式（ラジオボタン、プルダウン） 複数の選択肢から1つだけ選ぶ。

ex. 「情報」の授業は好きですか？

とてもそう思う／そう思う／ふつう／そう思わない／全くそう思わない

(2) 複数選択式（チェックボックス） 複数の選択肢から複数個選ぶ。

ex. あなたが好きなペンギンの種類を全て選んでください。

①マゼランペンギン ②アデリーペンギン ③ジェンツーペンギン

④ヒゲペンギン ⑤フンボルトペンギン ⑥コウテイペンギン

(3) 自由記述

【提出の方法】

(1) アンケートを1～2問作成し、友達に一度答えてもらう（これが重要、自分では問題点に気付けない）

(2) 問題がなければ、Classiに投稿する。

※共通して性別（男／女、ただし任意）、内進／外進、部活動（運動部／文化部／所属無し）は聞くので、質問としないでよい。

※Google Formの機能を利用する。

【今後の予定】

●説明 1組 9/1, 2組 9/14, 3組 9/11, 4組 9/11

●予備調査 1組 9/14, 2組 9/17, 3組 9/15, 4組 9/14

●アンケート作成めく 9/21(敬老の日)

●アンケート配布(Classi) 9/23(水)→作成1回目までに宿題として回答しておくこと

●作成1回目 1組 9/28, 2組 10/1, 3組 9/29, 4組 9/28

●発表1回目 11月第1週連休以降

Try2. アンケート項目の設定

1. ○○高校 ○○○○先生

仮説	○○○○
仮説の理由	あ あ
聞くべきこと	あ あ
結果 (スクショで挿入)	
結果・考察	
次のサイクルで扱いたいこと	

青森研修用アンケートフォーム

質問 回答 設定

青森研修用アンケートフォーム

フォームの説明

このフォームでは、すべての回答者からのメールが自動的に収集されます。 [設定を変更](#)

【武善-1】

☐ オプション1

【武善-2】

具体的なアンケート文と分析方法を考えてみましょう(&作成も一緒にやります)

※1人2～3設問まで

(例) アンケート文

①あなたは平日1日に、平均で何分くらい勉強をしていますか？(学校の授業以外、塾等も含む)。

_____分

②あなたは平日1日に、平均で何分くらい睡眠を取っていますか？

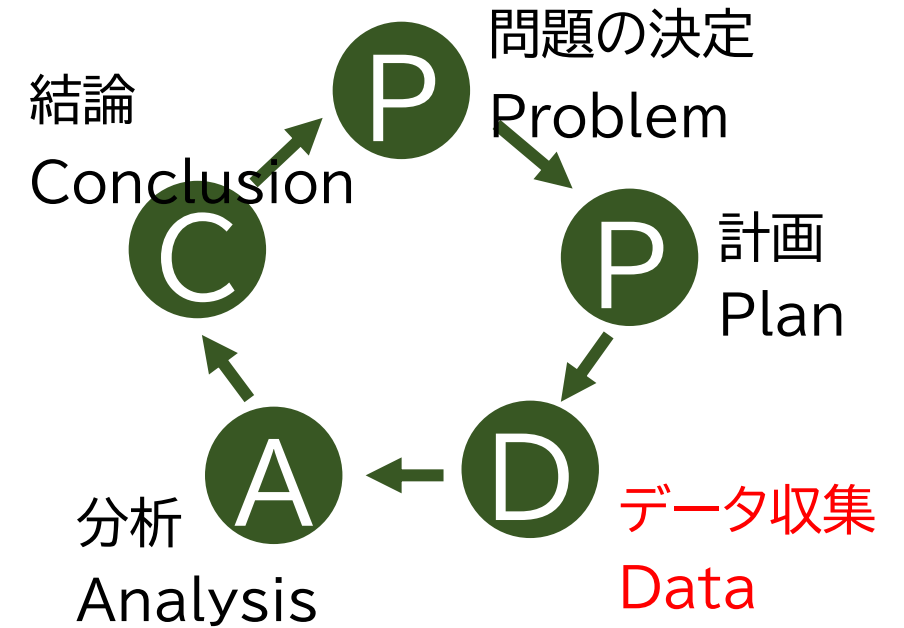
_____分

【分析方法】

①と②のデータを散布図に表して、相関係数を求めてみる。

ご回答をよろしくお願いいたします。

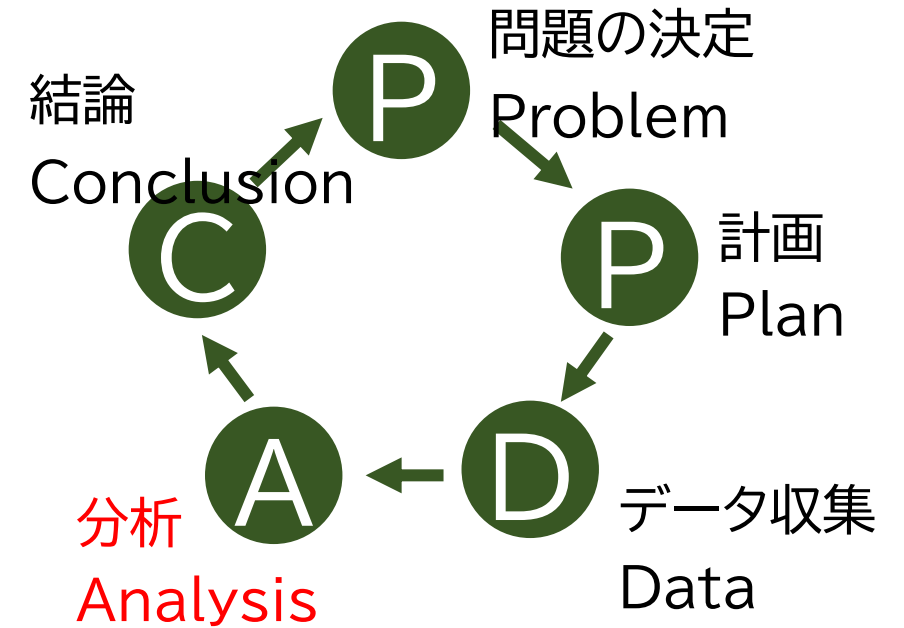
※回答後、ご休憩下さい※



Data データを収集し整理する

集まったデータを見てみましょう

- スプレッドシートを開き、「コピー」を作成。



Analysis データを分析する

例題と一緒に確認

- テスト質問
- テストデータ

相関係数(CORREL)

A.「自動車保有台数」と「人口」↔

B.「平均気温」と「平均積雪量」↔

C.「ピアノ保有率」と「降水量」↔

--

-1.0 ~ -0.7	-0.7 ~ -0.4	-0.4 ~ -0.2	-0.2 ~ 0.2	0.2 ~ 0.4	0.4 ~ 0.7	0.7 ~ 1.0
強い負の相関	負の相関	弱い負の相関	相関なし	弱い正の相関	正の相関	強い正の相関

-1 ~ 1の値をとる。

Try3.

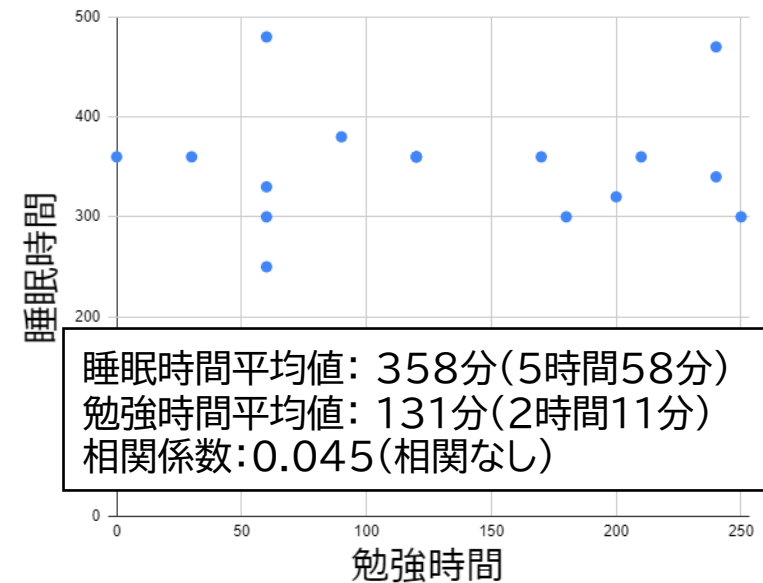
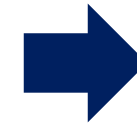
1. ○○高校 ○○○○先生

仮説	○○○○
仮説の理由	ああ
聞くべきこと	ああ
結果・考察	
結果 (スクショで挿入)	
次のサイクルで扱いたいこと	

収集した結果を、グラフ等で表してみよう。

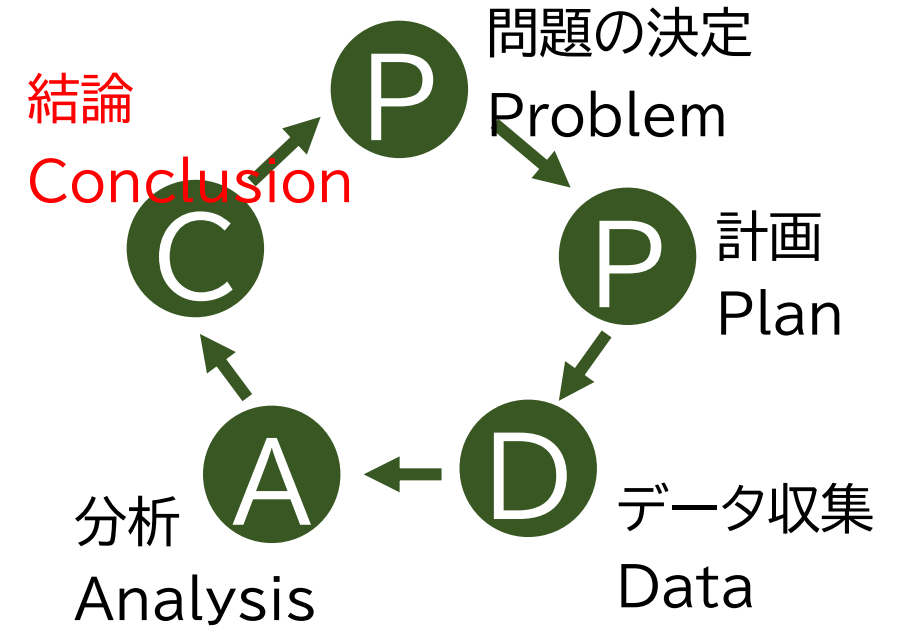
<例>

回答 №	あなたは平日1日 に、平均でどれくら い勉強をしています か？（回答単位： 分）（学校の授業以 外、塾等も含む）	あなたは平日1日 に、平均でどれくら い睡眠を取っていま すか？（回答単位： 分）
1	250	300
2	60	480
3	180	300
4	1	5
5	60	250
6	120	360
7	120	360
8	0	360
9	240	470
10	240	340
11	120	360
12	90	380
13	60	330
14	170	360
15	30	360
16	200	320
17	210	360
18	150	450



ただし、これらの分析はもう生成AIで出来る

- スプレッドシートの関数を知らなくても、AIに「このデータの標準偏差を計算して」と頼むだけ。
- ただ、分析のイメージや意味合いを知らないとそもそも命令できない&解釈できない。
- 生成AI → 「手綱を握る力」



Conclusion 仮説の考察

どうしてこのような結果になったのか？

- 考察

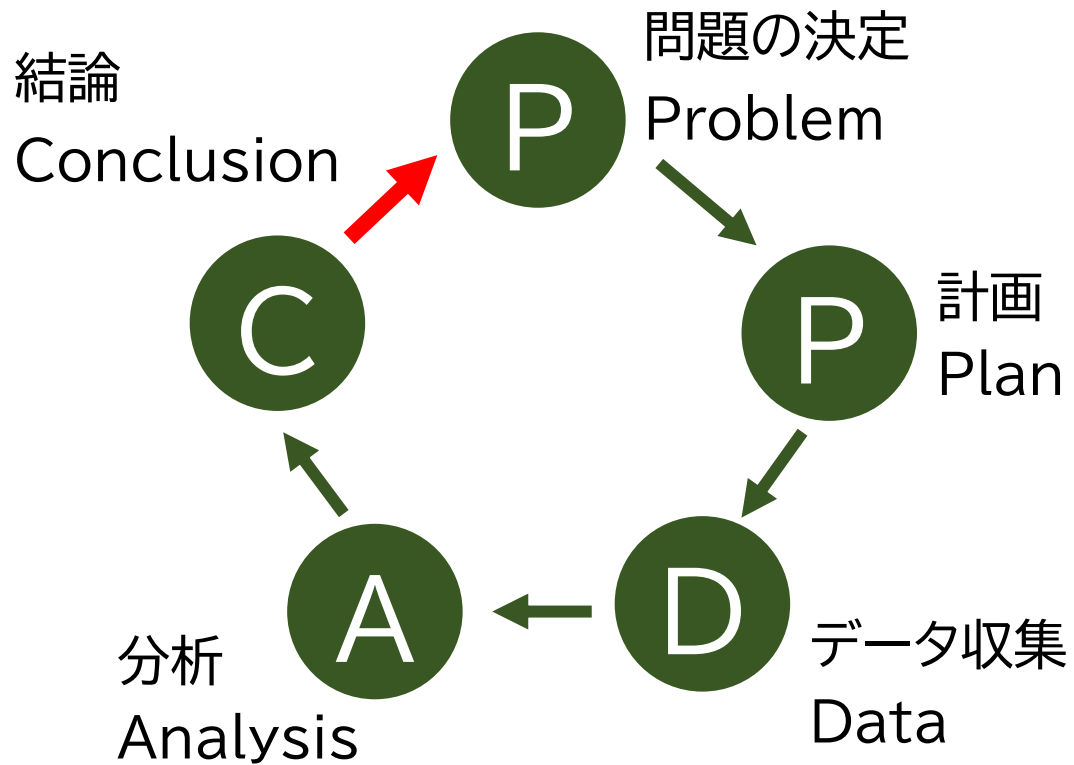
- 多角的に結果を見て、考えたこと
 - アンケート結果が「なぜそうなったのか」
 - 仮説と「なぜ違ったのか」
 - 結果を受けて「どう考えるのか」 など

- 結論

- 結果と考察をもとにわかったこと



PPDACサイクル



- 次のサイクルに繋がるように終わらせることが大切。
- 仮説が間違っていたなら、他に考えられる仮説は？
- その仮説を証明するのに、必要なデータは？分析方法は？

Try4.

1. ○○高校 ○○○○先生

仮説	○○○○
仮説の理由	ああ
聞くべきこと	ああ

結果
(スクショで挿入)

結果・考察

次のサイクルで扱いたいこと

結果を見て、
考えたことを書いてみよう！

<例>

睡眠時間と学習時間に相関がなかったのは、なぜ？

- ・部活動や通学時間など他に使っている時間による要因によって違うから、仮説と違った。
- ・今後、「部活動」「通学時間」など、他の要因に関してもアンケート調査を実施してみる。

勉強時間が長くても、睡眠時間をしっかり取れている人がいる。生活の上で、どのような工夫をしているか、調べてみたい。

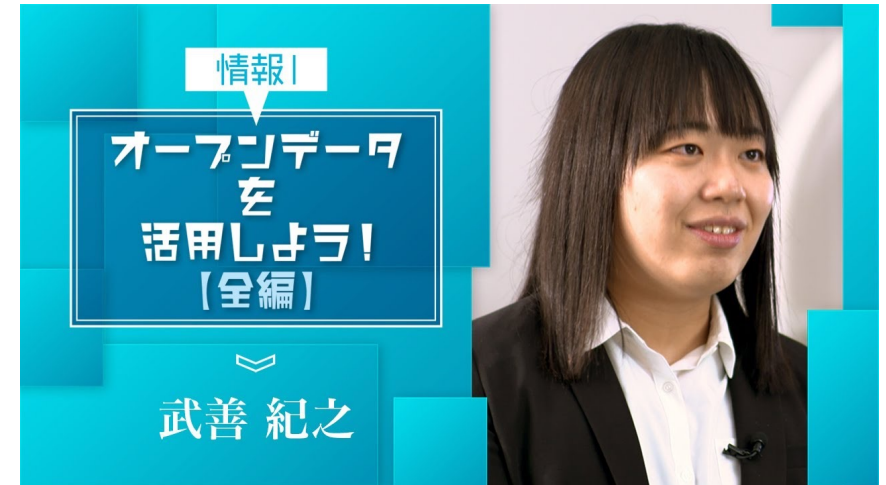
終わり

(参考) オープンデータを活用する場合のサイクル

文科省の動画では2パターン実践

①e-Statのデータから分析

②地球規模のデータを活用した問題解決

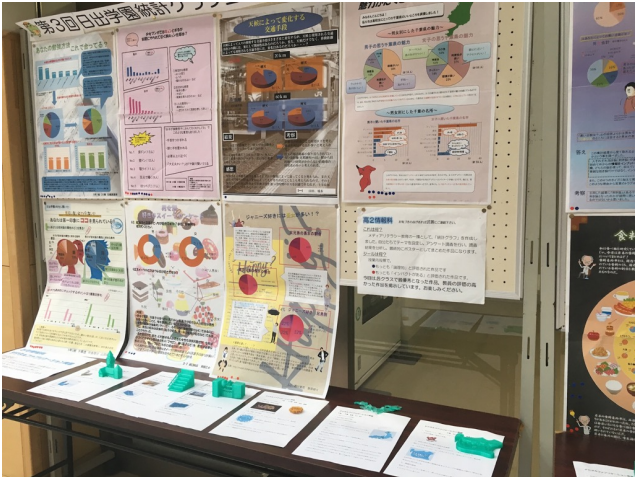


3. 実際の指導例

統計グラフをやっていた頃の例

2015～2019 統計グラフポスターの作成(チーム)

1	問題解決実習の概要説明
2	テーマ設定＋予備調査
3	グラフの作り方+調査①
4	調査②
5	調査③orアンケート作成
6	まとめ①orアンケート分析
7	まとめ②
8	まとめ③
9	発表会

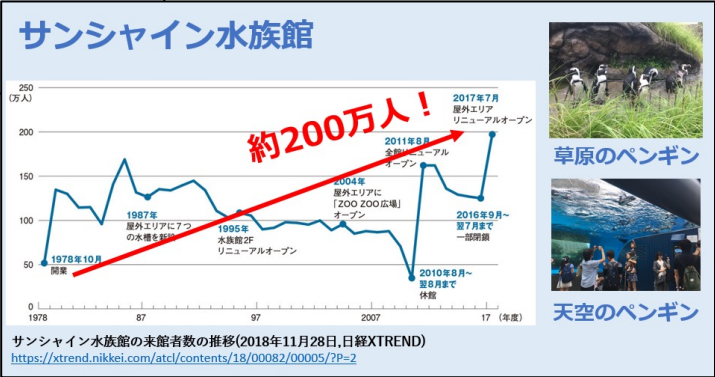
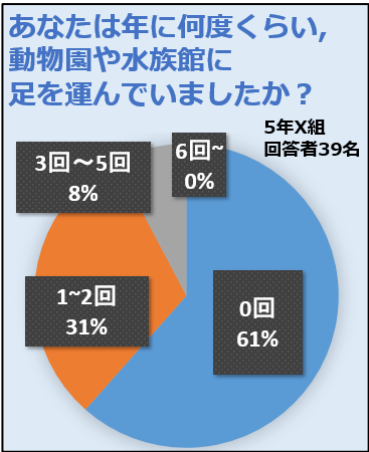


2020「統計調査を含んだ」という条件でプレゼン実習(個人)

1	問題解決実習の説明
2	予備調査, アンケートの作成 準備(1人2設問まで)
宿題	①アンケート作成 ⇒ 武善 が集計 & 全体へ配布
	②アンケート回答 (Google Form)
3~5	スライドの作成 (アンケート 回答結果は1コマ目に配布)
6~11	1コマ当たり8人が発表

そんなペンギンも...

10 / 18



現在、「とどラン」をやっている例

2024～ 探究と役割分担(PPDACの基礎のみを学ぶ)

1	とどランの説明・作成
2	作成・ギャラリーウォーク型 発表
3	発展的な分析(本日午後の 内容)
(4)	オープンデータを眺める



日出学園中学校・高等学校「【情報科】『こんな関係性がありそう!』データサイエンス入門」日出学園、2026年2月22日、閲覧日:2026年9月21日、<https://www.hinode.ed.jp/high/blog-h/36657/>

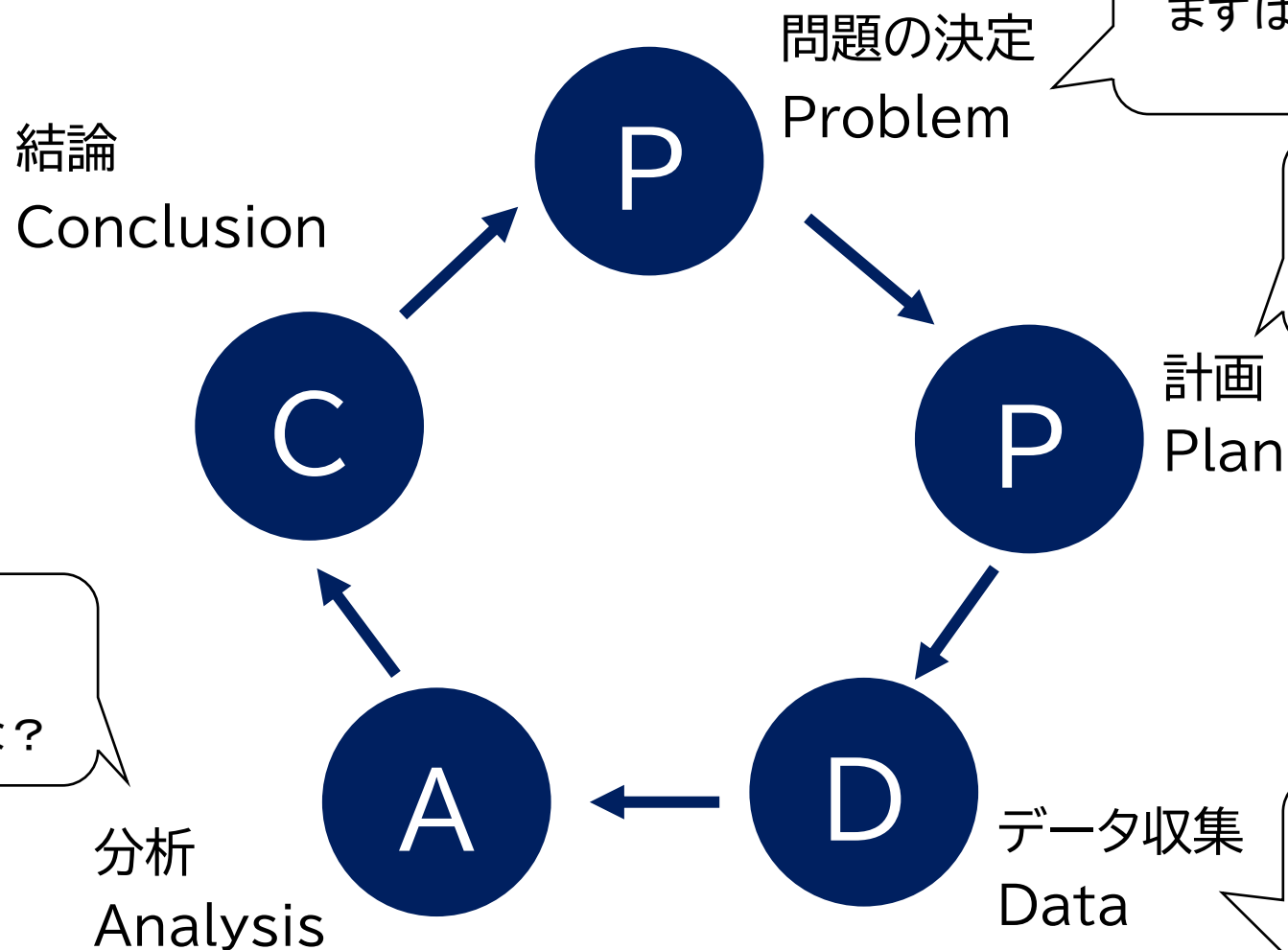
この学びがうまく「総合的な探究の時間」に繋がって欲しい、という理想。

ちなみに共通テスト「情報Ⅰ」第4問は題材の宝庫

- これを数本解いたり、眺めたりしても面白いです。

ここまでのまとめ

「頭」の中にPPDACのイメージがあることが大事



課題が見つからなければ、
まずは「なんでもアンケート」を
やってみよう

アンケートじゃなくて、
こんなオープンデータも
あるよ

このデータなら、
〇〇な分析も
試せるんじゃないかな？

その設問だと
使えるデータにならないよ

探究担当者と話して感じたこと

- 「問い立て」がやはり課題
-

- 早い段階でサイクル全体を見せることが、テーマそのものも考えやすくする？
- 「生成AIについて研究したい」
↓
次に何をすればよいか分からない……
- 何を知りたい？ (Problem)
→ 何を集める？ どう見る？ どう分析する？ (Plan)
→ 分かったら次に何をする？
- 研究計画≡迷子にならないための仮の地図
- 学期完結程度の学習であれば仮説発見でもいいけるが、年間の時間軸になるとある程度の仮説・目的がないと厳しい。

後半では……

- 指導する上で、実際に困ること（分析以外の課題）
- 統計的に持っているの良い「 $+\alpha$ 」（解析はAIとコンピュータがやってくれます）

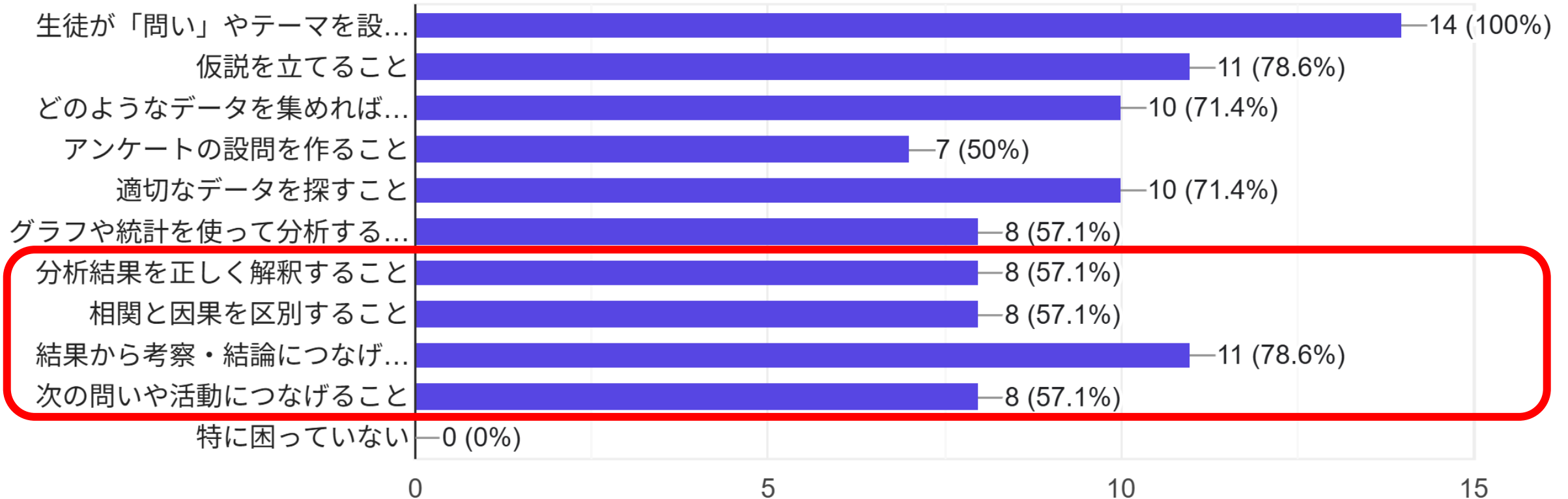
等を話します。

4.研究上の落とし穴

(再掲)事前アンケートの結果より

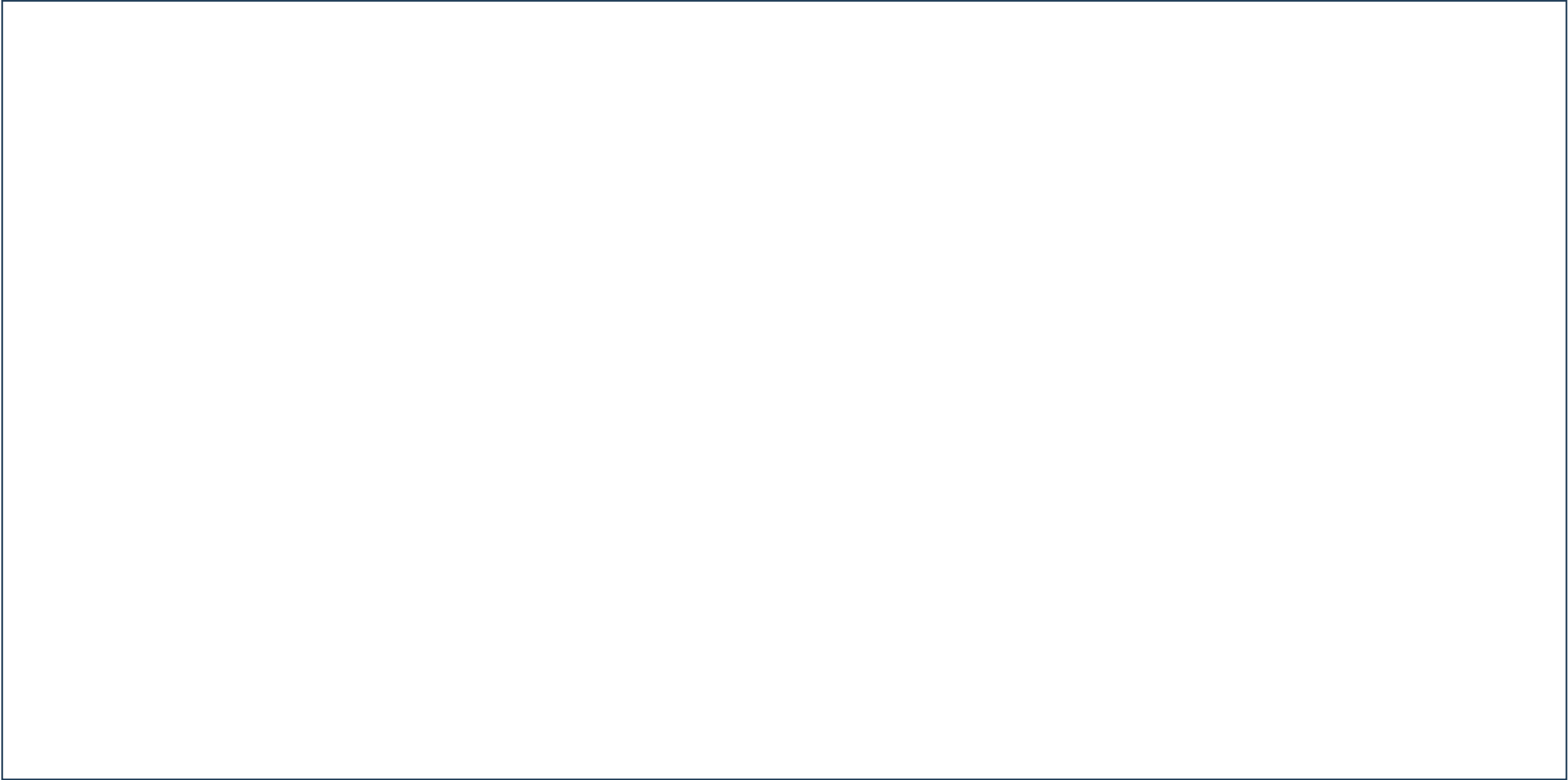
10. 生徒の探究活動を指導する際、難しい・困ると感じることをすべて選んでください。

14件の回答



- ・後半で扱う箇所を中心。
- ・ファシリテート(伴走)するためのコツ。

①アンケート調査は本当に気軽？



参考資料
東京書籍
「新編情報Ⅰ」
紙面より

アンケートの「品質」と「配慮」

「アンケート調査」は、
取った後に「あれも聞けば良
かった...」「これも聞くべき
だった...」が多い。

東京書籍「新編情報Ⅰ」
紙面より

(再掲)設計に関する技術的な問題

【答えにくいアンケートにしないコツ】 ←

(1) ことばづかい・文章表現(wording) ←

ex. ヒゲペンギンとマカロニペンギンのどちらが好きですか？（みんなはペンギンに詳しくない） ←

ex. 「宿題は少ないが、いつも冷たい先生」と「いつも冷たいが、宿題は少ない先生」 ←

(2) あいまいな表現 ←

ex. ペンギンはよい生きものであると思いますか？（よい生き物って・・・？） ←

(3) ダブル・バーレル質問（二重質問）(double-barreled question) ←

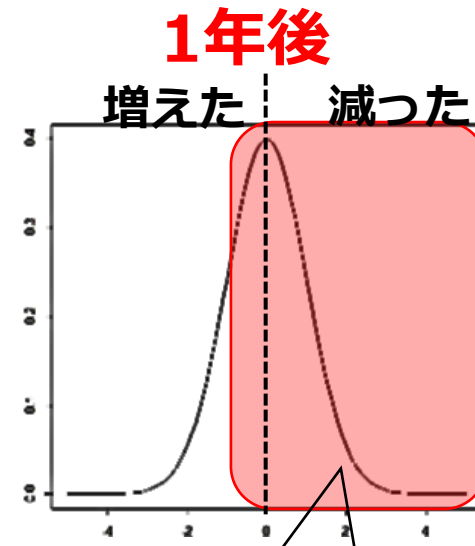
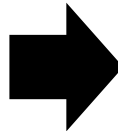
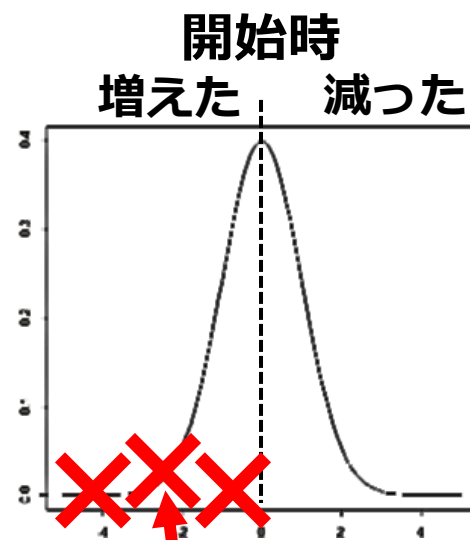
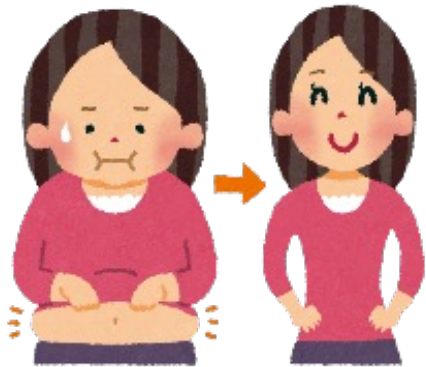
ex. 「あなたは日本食や日本のアニメが好きですか？」（日本食好きだけどアニメ嫌いな人は？） ←

避けるべき「バイアス(偏り)」

- ・ 設問文はバイアスを極力、除去するためにある
- ・ もっと言えば、誰に聞くかの時点で「バイアス」は掛かる。
- ・ 例えば、生徒がよくやる「SNSアンケート」はあまり意味がない。

大学より
日出が大好きな
卒業生

1年後調査で
95%が
痩せていた
奇跡のジム！



95%が
痩せた！



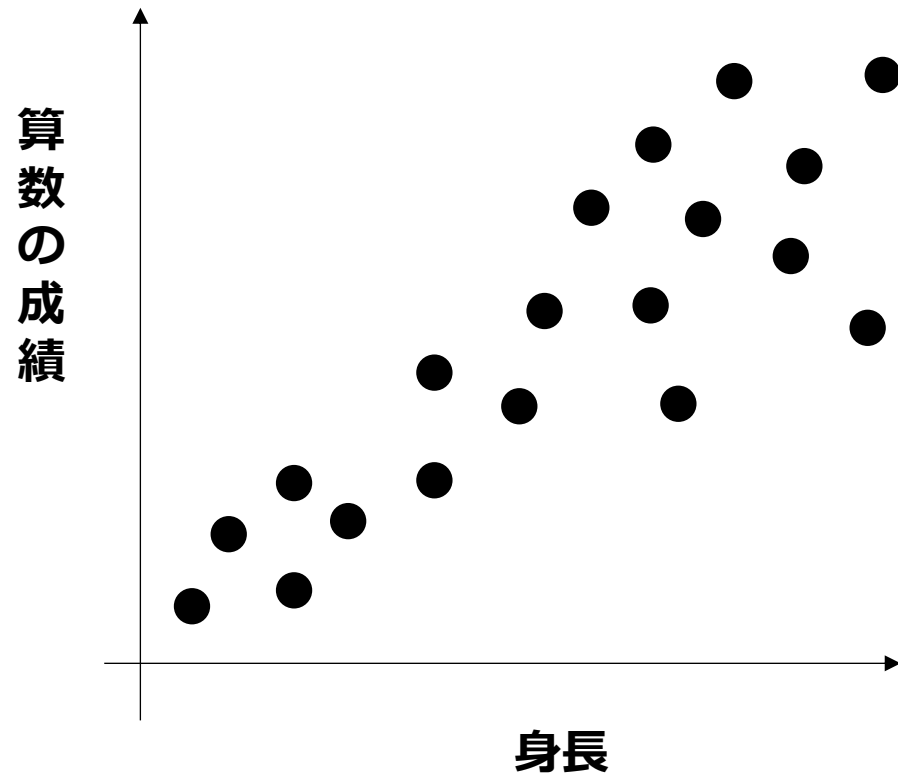
行くの面倒だな～



効果ないや。
他のダイエットをやる～

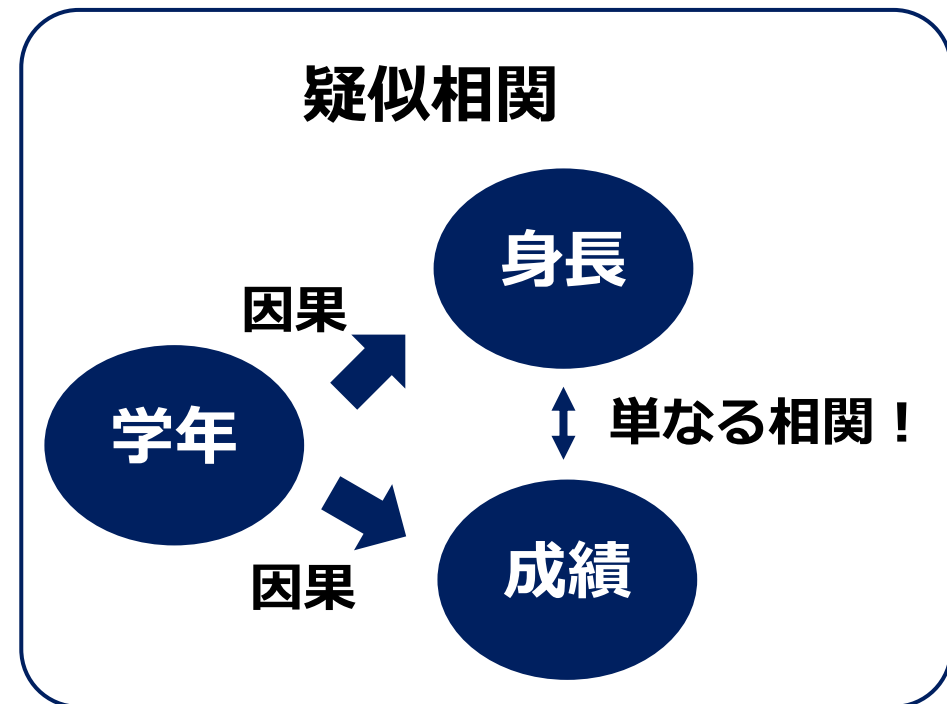
②その分析、本当に正しい？

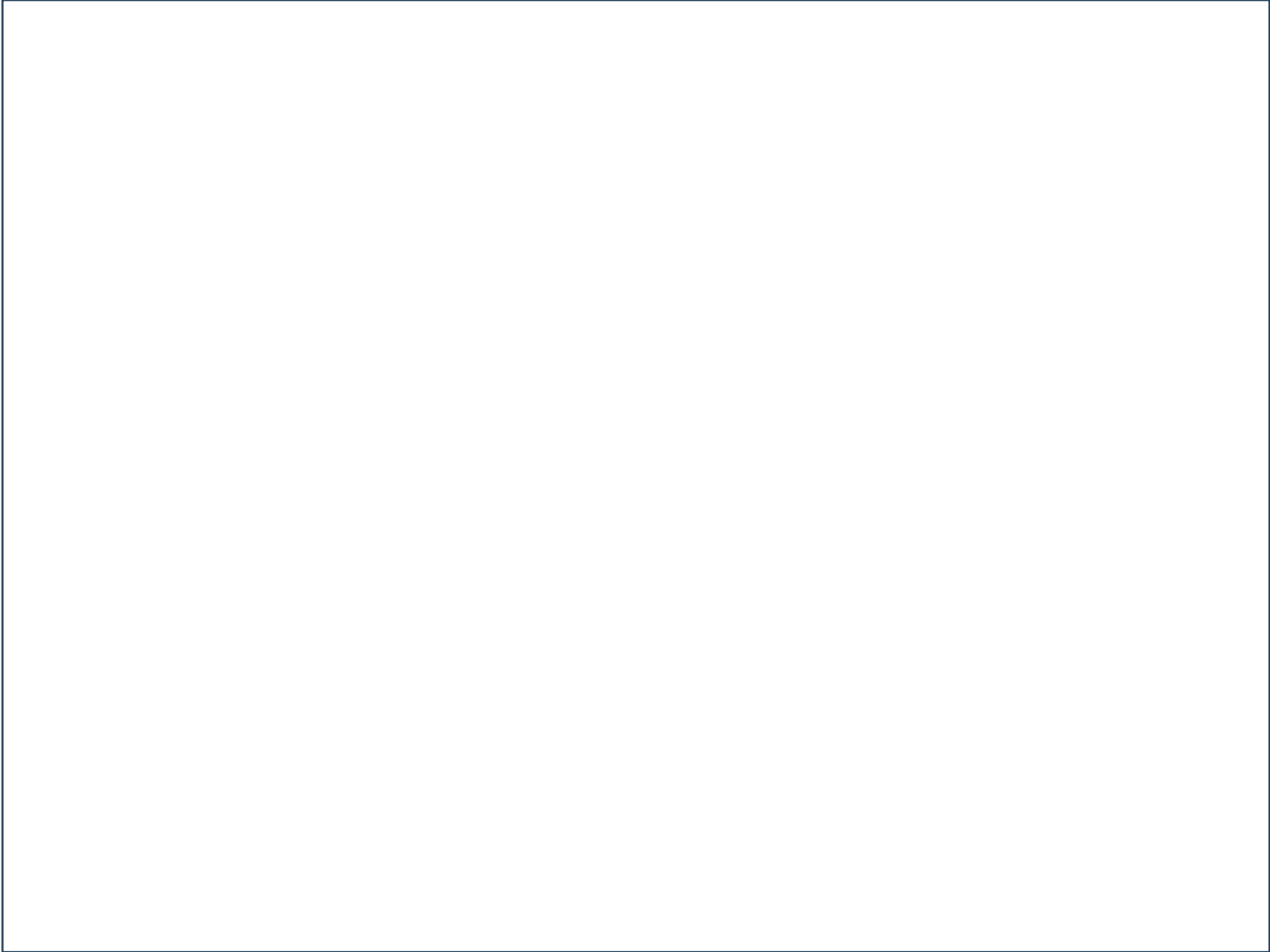
相関・因果の関係



- 小学生32人を無作為に抽出
- 算数のテスト(縦軸)、身長(横軸)
- 身長の高い子供の方が、算数の点数が高い！

参考
[統計的消去で擬似相関を見抜こう！](http://hoxo-m.hatenablog.com/entry/20130711/p1)
<http://hoxo-m.hatenablog.com/entry/20130711/p1>





因果関係の3基準

1. 共変関係(相関関係があること)
2. 時間的順序関係
3. 第3変数(交絡因子)の排除

生徒の「探究」では……？

しかし、武善はしつこかった

第12回全国高等学校
情報教育研究大会
(和歌山大会)

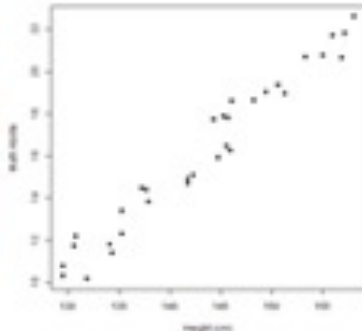


本当に「運動部に入ると、
運動が好きになる」？

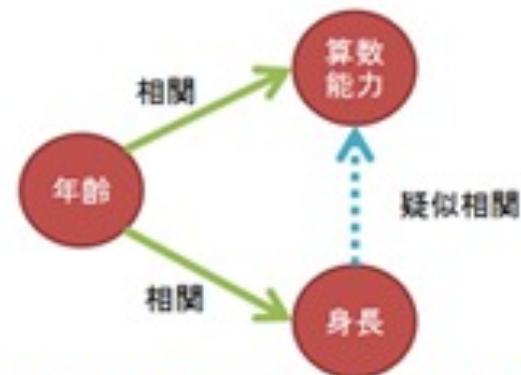
運動が好き**だから**、運動部に入ったんじゃないの？

親がスポーツに熱心**だから**、運動部に入部させて、
子供も運動を好きになったのでは？

算数の成績



身長



<http://hoxo-m.hatenablog.com/entry/20130711/p1>

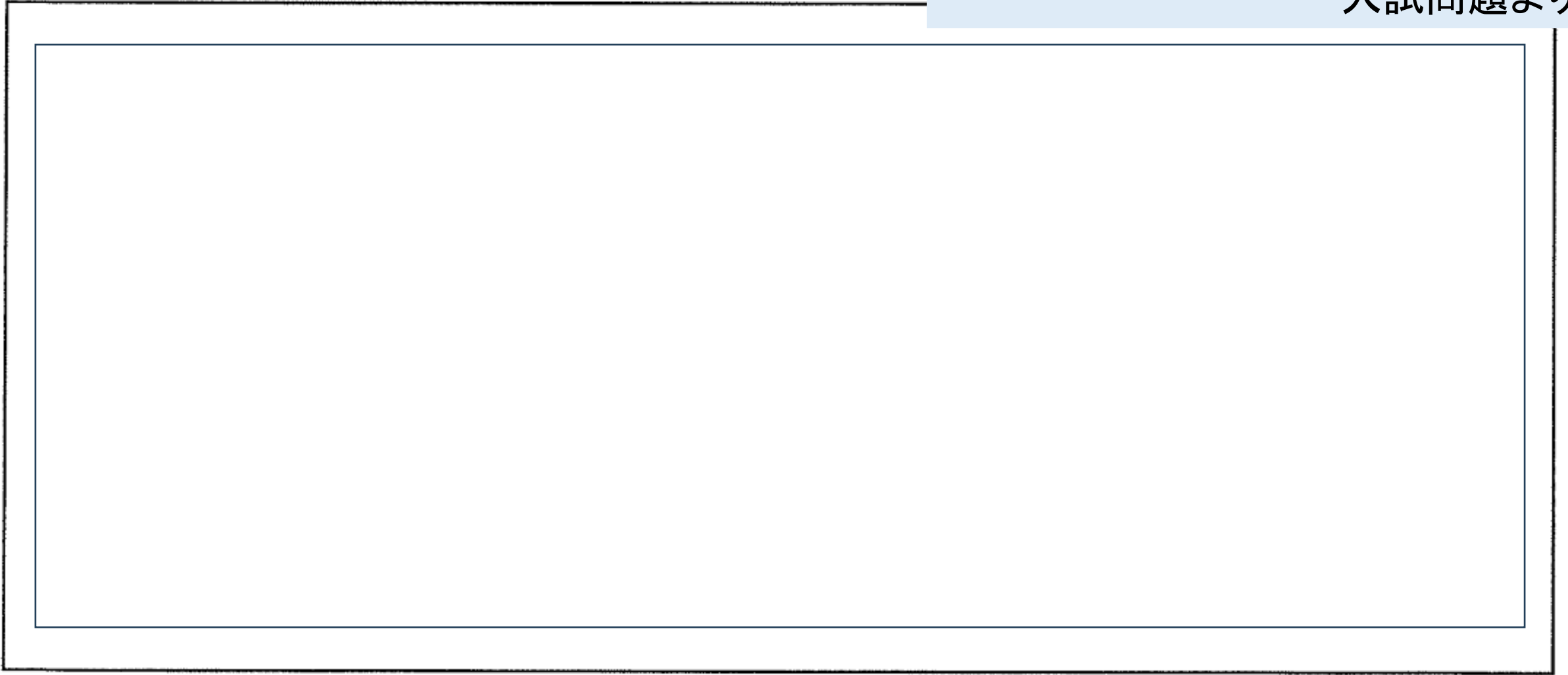
武善紀之「高校生が楽しく学べる仮説
検定と相関・因果—シミュレーションから始まる『情報科』統計教育」『キミのミ
ライ発見』河合塾、閲覧日:2026年9月
21日、<https://www.wakuwaku-catch.net/jirei19136/>

大学入試問題

慶應義塾大学 総合政策学部2017
入試問題より

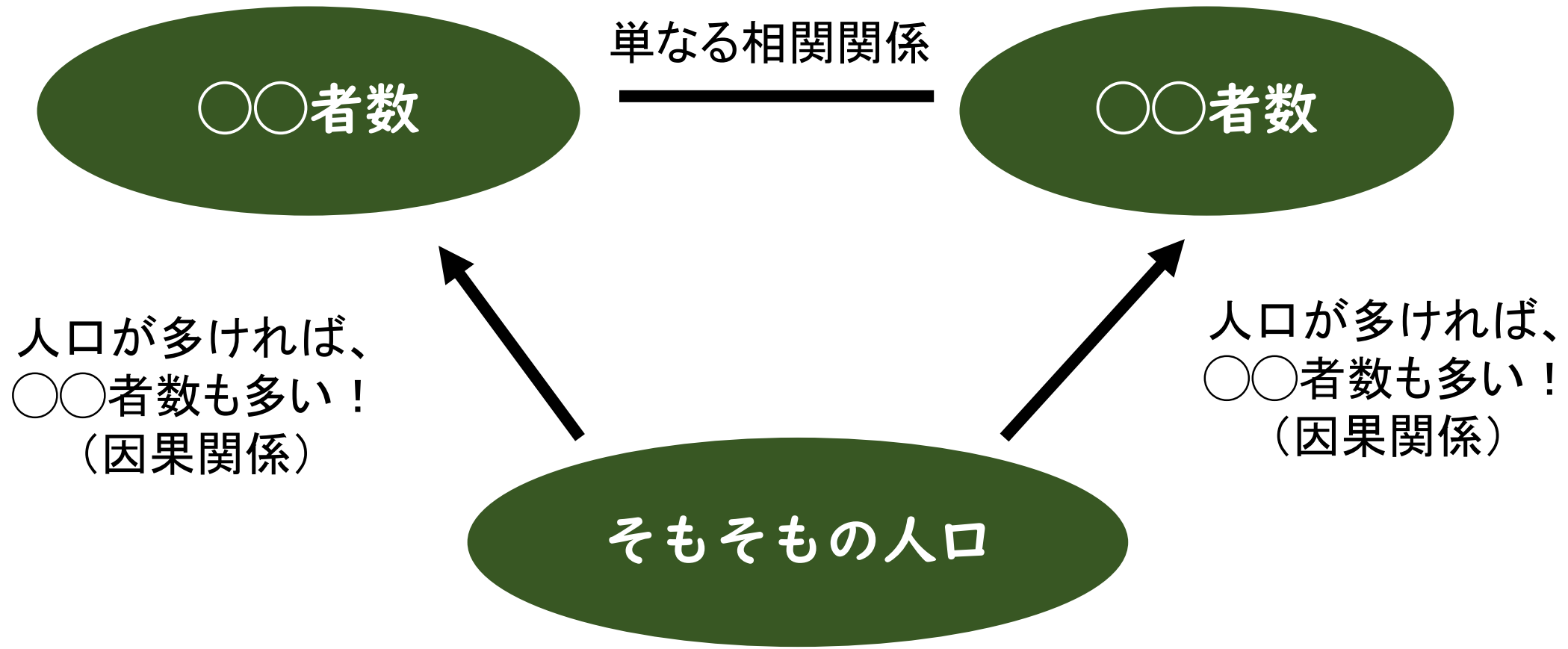
こんな関係性があるのでは？

慶應義塾大学 総合政策学部2017
入試問題より



「とどラン」で分析するとよく起きること

この関係性に陥っている人が多いそう。



Q.このようなデータを分析するためには？

この辺りで生徒が感じること

- 重箱の隅をつつくようなことばかり
- せっかく結果を出せて楽しんでいるところに、水を差された…。

統計をやると、性格が悪くなる？

「人間の感覚」と「統計的な考え方」

「雨乞いをすると雨が降る」

- ・ 雨乞いしたら、雨が降るんだ！
- ・ 雨乞いをしないときには、雨は降っていない！も観測する必要あり
- ・ ところが、確証バイアスによって左上にばかり注目してしまう

雨乞い／雨	降る	降らない
する	56	
しない		

「人間の感覚」と「統計的な考え方」

「雨乞いをすると雨が降る」

- ・ 雨乞いしたら、雨が降るんだ！
- ・ 雨乞いをしないときには、雨は降っていない！も観測する必要あり
- ・ ところが、確証バイアスによって左上にばかり注目してしまう

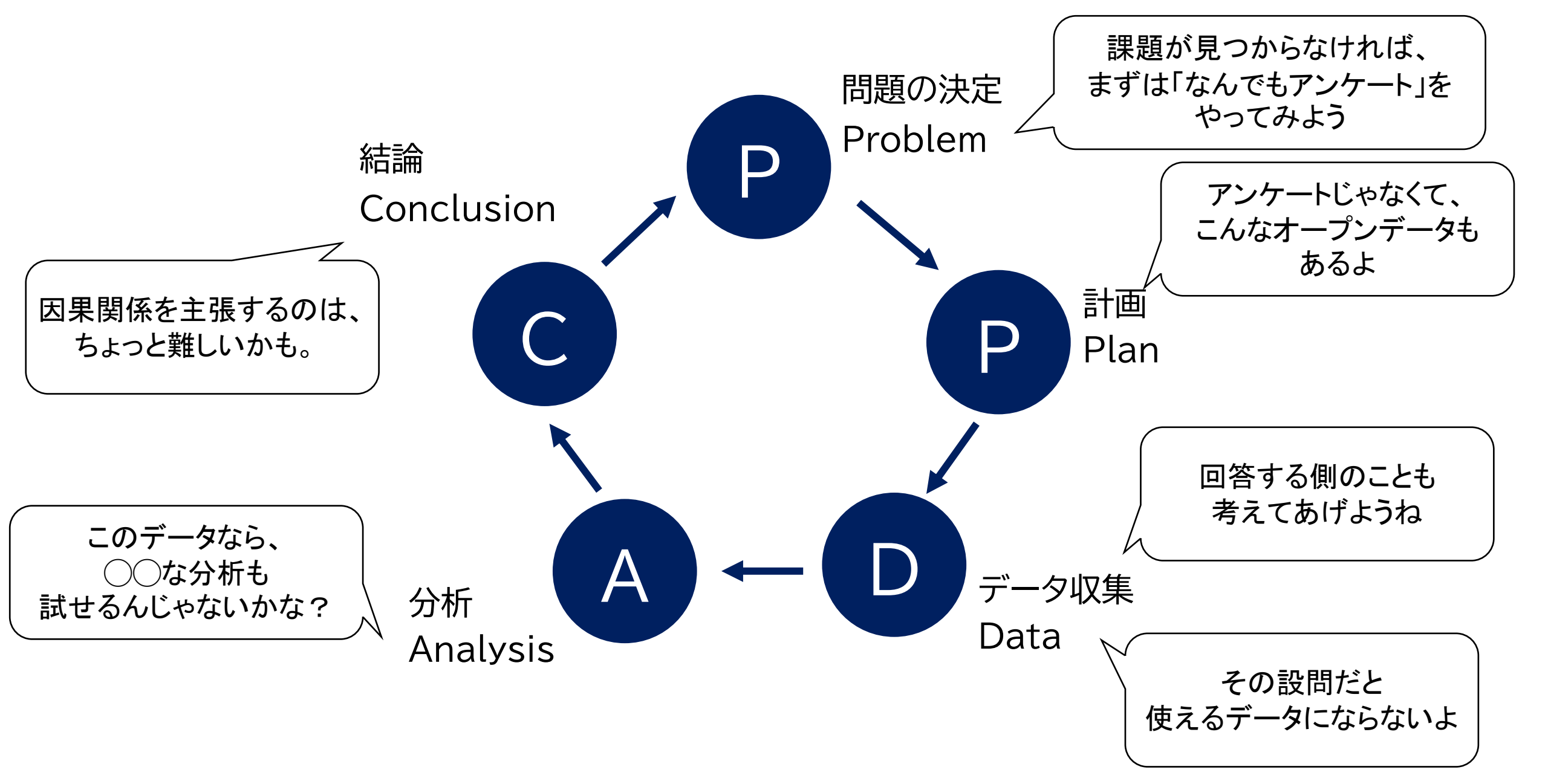
雨乞い／雨	降る	降らない
する	56	10
しない	21	62

「新しい見方・考え方」を提供してくれる

ここで……

- ・ 自分の午前中の分析を振り返ってみましょう。

これらが「頭」の中にイメージだけあることが大事



5. ステップアップした分析

「こんな見方もあるのか」で十分です！

①数学 I の学びをつなごう

ステップアップの分析要素

(数学Ⅰまでの範囲でも他に……)

- 代表値 (平均値以外)
- 箱ひげ図
- 標準偏差

数学の先生はご存知のことばかりになりますが…。

ステップアップの分析要素

平均値だけでいいの？

次のデータは、5人の生徒の情報の試験の成績（100点満点）である。平均点を計算せよ。

15, 20, 95, 20, 25

(答え)

$$15 + 20 + 95 + 20 + 25 = 175$$

$$175 \div 5 = 35 \text{ (点)}$$

この表示は適切？

35分以下の人が4人もいるぞ？！
データに極端な数値があるときは、
「中央値」を考える。
今回の場合は、25分

ステップアップの分析要素

「中央値」と「平均値」

パーソルキャリア株式会社「正社員の年収中央値
は？年齢別・都道府県別にも解説 平均年収ランキン
グ【2022年版】」doda、2023年2月27日更新、閲覧日：
2026年9月21日、
<https://doda.jp/guide/heikin/median/2022/>

ステップアップの分析要素

平均値が同じならOK?

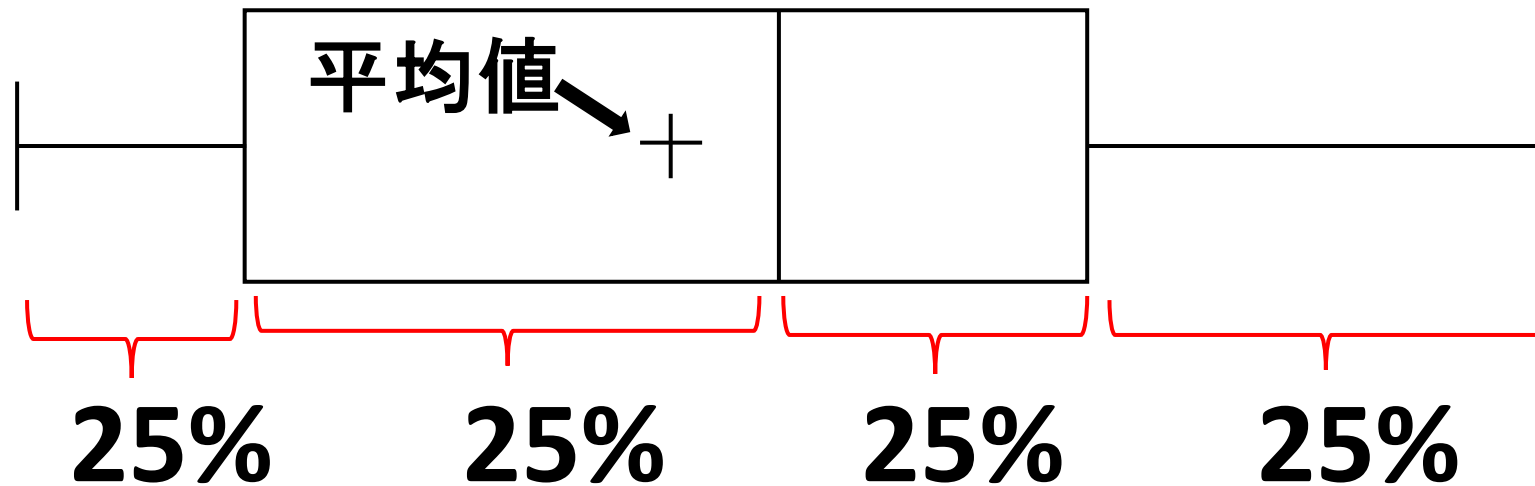
平均点60点の情報

平均点60点の数学

ステップアップの分析要素

「箱ひげ図」で散らばりを見る(Excelだと作れる)

箱ひげ図のポイントはこれだけ！



○
↑
外れ値

ステップアップの分析要素

箱ひげ図を作ると？

<https://xn--48s96ub7b0z5f.net/hakohige-hisutoguramu/>

平均点60点の情報 平均点60点の数学



ステップアップの分析要素

散らばりを数値で表現 → 標準偏差

平均点60点の情報

平均点60点の数学

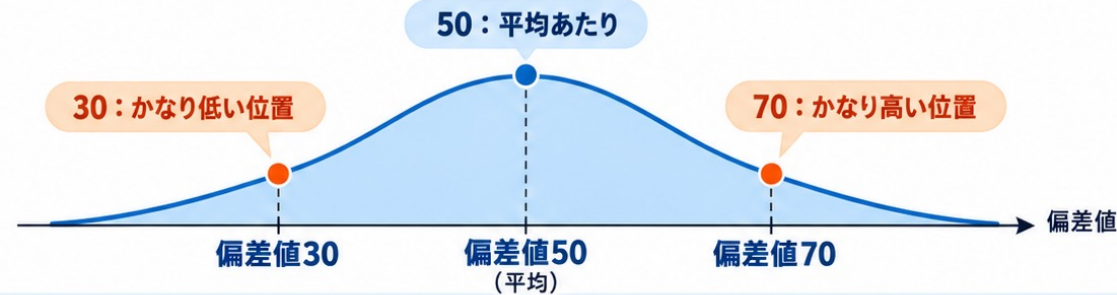
標準偏差

ステップアップの分析要素

ちなみに……偏差値は「平均」と「標準偏差」で決まる

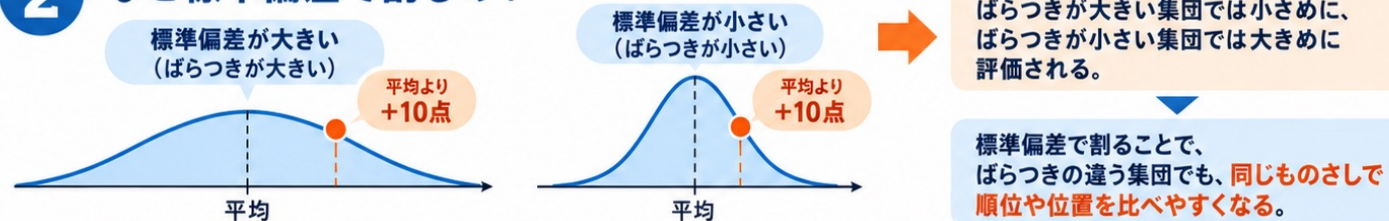
$$\text{偏差値} = (\text{得点} - \text{平均}) \div \text{標準偏差} \times 10 + 50$$

1 偏差値は“集団の中でどの位置にいるか”を表すものさし



偏差値は、点数そのものではなく“平均との差を、ばらつきも考えて表した値”。

2 なぜ標準偏差で割るの？



同じ +10 点でも、ばらつきが大きい集団では小さめに、ばらつきが小さい集団では大きめに評価される。

標準偏差で割ることで、ばらつきの違う集団でも、同じものさしで順位や位置を比べやすくなる。

3 つまり… 偏差値は、平均を50として、集団の中で自分がどのあたりにいるかを表した数値。

標準偏差で割ることで、広い分布も狭い分布も、同じ尺度で比べられる。

ステップアップの分析要素

分析の練習

A large, empty rectangular box with a thin black border, intended for writing or drawing during the analysis practice session.

ステップアップの分析要素

回帰分析

2つの事柄の関係性を数式の形で表す手法

ステップアップの分析要素

回帰分析

$$y = 6.4205x + 370.87$$

飲料販売量

平均気温

気温が1℃あがるごとに、6.4本飲料が売れる。

例. 平均気温が34℃だったら……

$$\begin{aligned} Y &= 6.4205 \times (34) + 370.87 \\ &\div 589 \text{ (本)} \end{aligned}$$

②その結果、偶然じゃない？
-仮説検定-

「仮説検定」の利用場面

特進クラス
72点

進学クラス
70点

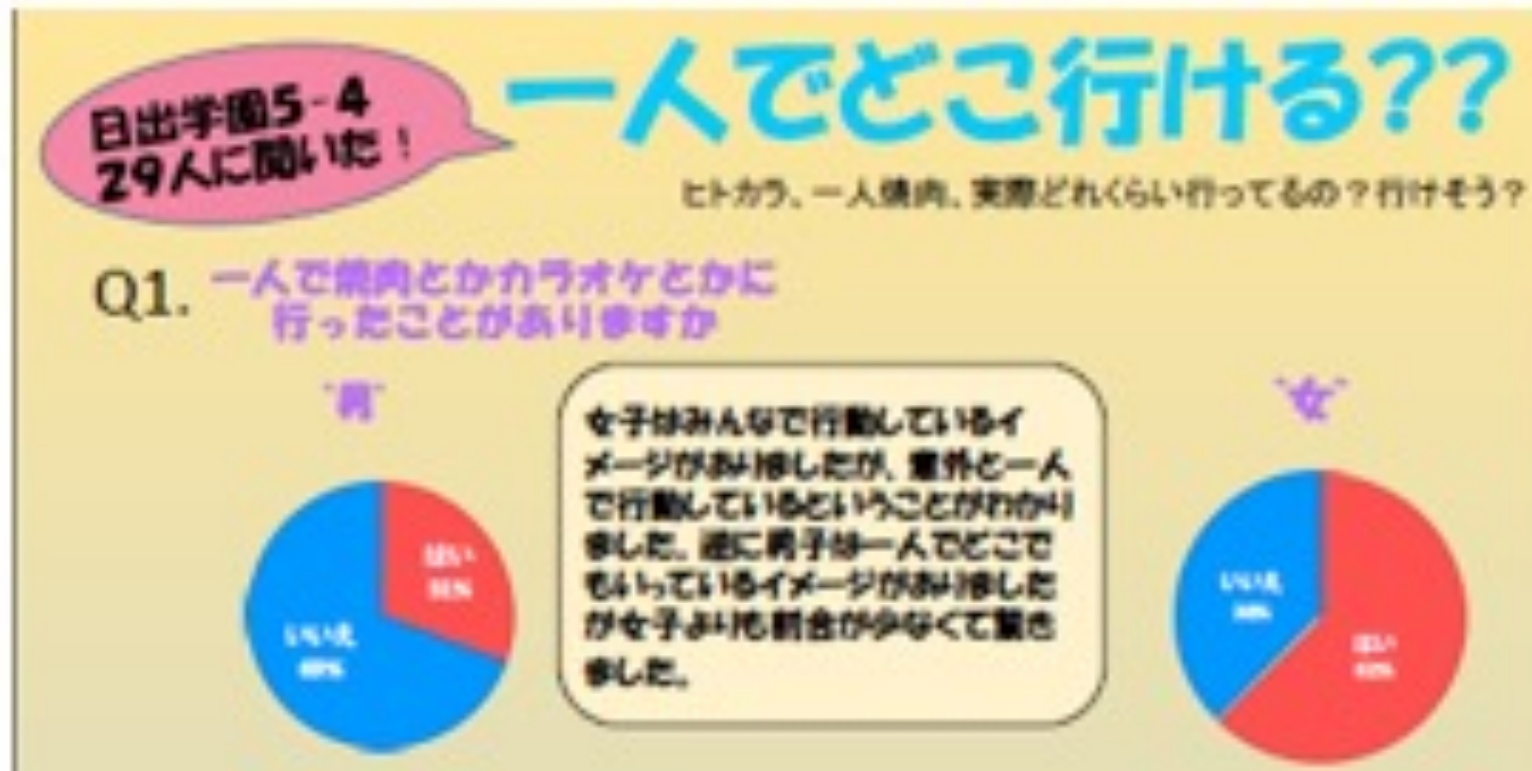
進学クラスより、
特進クラスの方が
優秀です！



え？ 本当？ たまたまじゃないの？

偶然 vs 必然(要因) のバトル

「仮説検定」の利用場面



今までの統計グラフ作成では、
偶然性を排除できていない

仮説検定の手順

Step1. 帰無仮説を設定。

A組とB組の点数には差がない (偶然生じたものだ)

Step2. いろいろ計算する。

<http://sys-assist.co.jp/archives/1781>

偶然にしては出来すぎている、おかしいぞ

A組	B組
72	47
78	67
65	70
42	89
67	56

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Step3. 帰無仮説を棄却。対立仮説を採択。

仮説が間違っていたに違いない。

つまり、**A組とB組の点数には差がある** [対立仮説]

SBIを用いた「仮説検定」の説明

Simulation-based Inference(SBI)

シミュレーションに基づく推論

- ・ 諸外国の入門統計教育の主流
- ・ informalな統計教育

cf. formalな統計教育（数学 B の内容的な）
確率変数や確率密度関数が必要

（出典）

Jimmy A Doi,

Active Learning Lectures for Statistical Understanding
(推測統計の概念を理解するためのアクティブラーニング授業),

第6回情報教育研究会 in 江戸川大学, July 2018

仮説検定はたったの3Step

1. 仮説を立てる。
2. 実験をして証拠を集める。
3. 証拠に基づいて結論を出す。

コイン投げはインチキか？

仮説 コインは公正

実験 50回このコインを投げる

予測 25回は表が出る

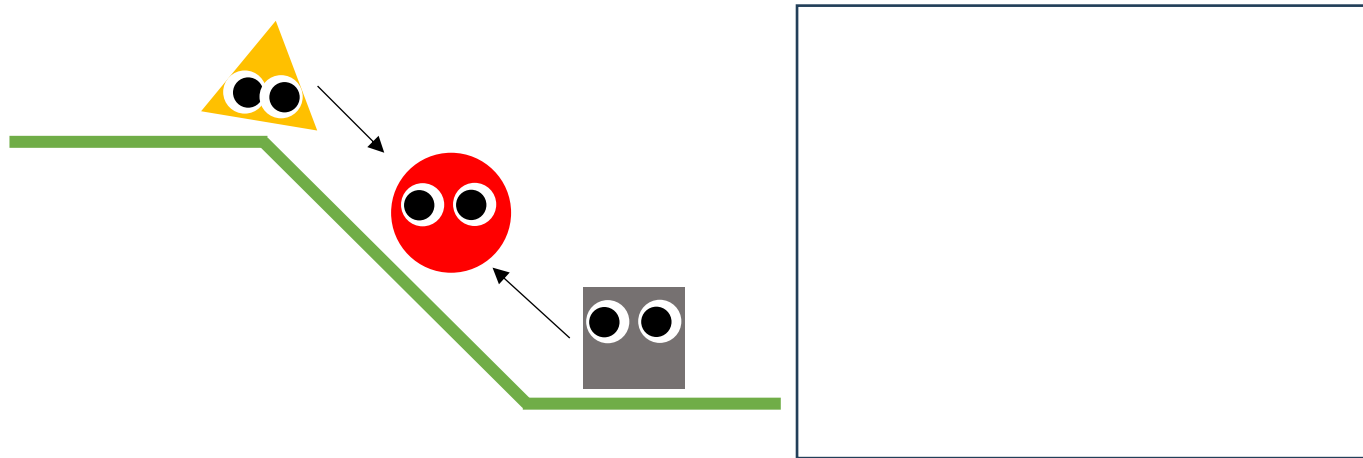


表の数	予想の範疇？	仮説は
24		
22		
2		

仮説検定はたったこれだけ！

Helper and Hinderer

- <https://www.youtube.com/watch?v=anCaGBsBOxM>
- Helper versus Hinderer
(赤ちゃんは善悪がわかるか?)

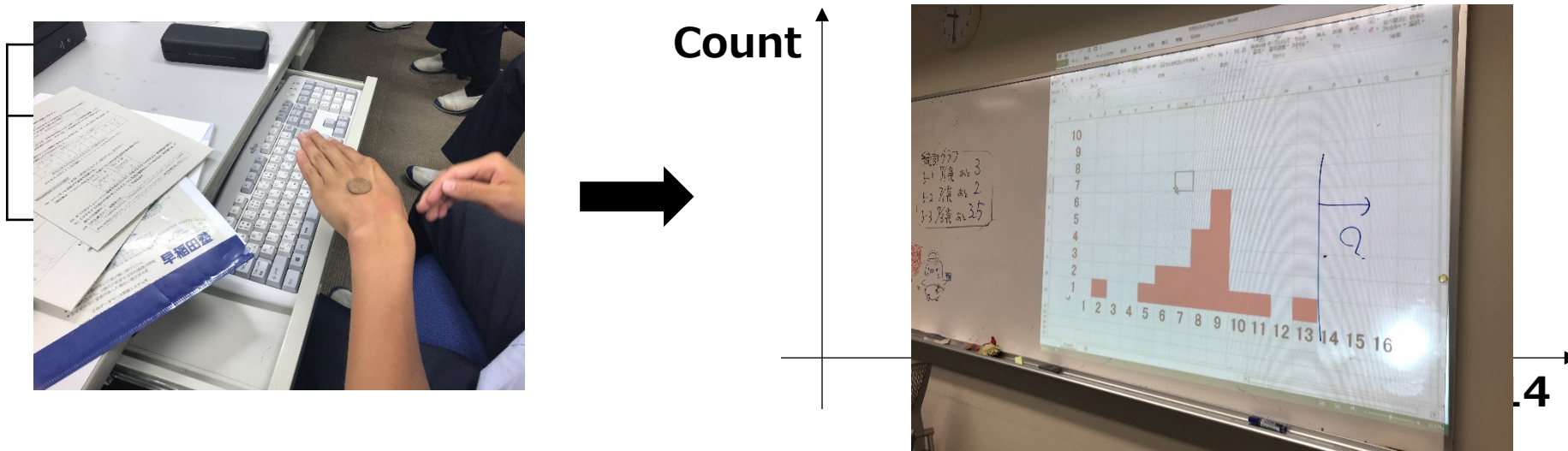


この実験を
シミュレーション
してみよう！



 Helper Selected	 Hinderer Selected	Total
14	2	16

Simulation1



激レア（偶然ではありえないことが起こった）

14回はあり得なかった（偶然じゃない！）

➡赤ちゃんには好みがある！



Helper Selected	Hinderer Selected	Total
14	2	16



Simulation2

PCを用いて、大規模シミュレーション

<http://www.rossmanchance.com/applets/OneProp/OneProp.htm>

Probability of heads:

Number of tosses:

Number of repetitions:

☐ Animate

☒ Number of heads

☐ Proportion of heads

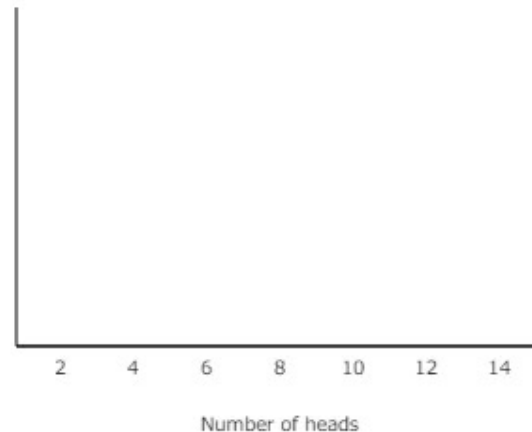
As extreme as

☐ Two-sided

☐ Exact Binomial

☐ Normal Approximation

☐ Summary Stats



(出典)

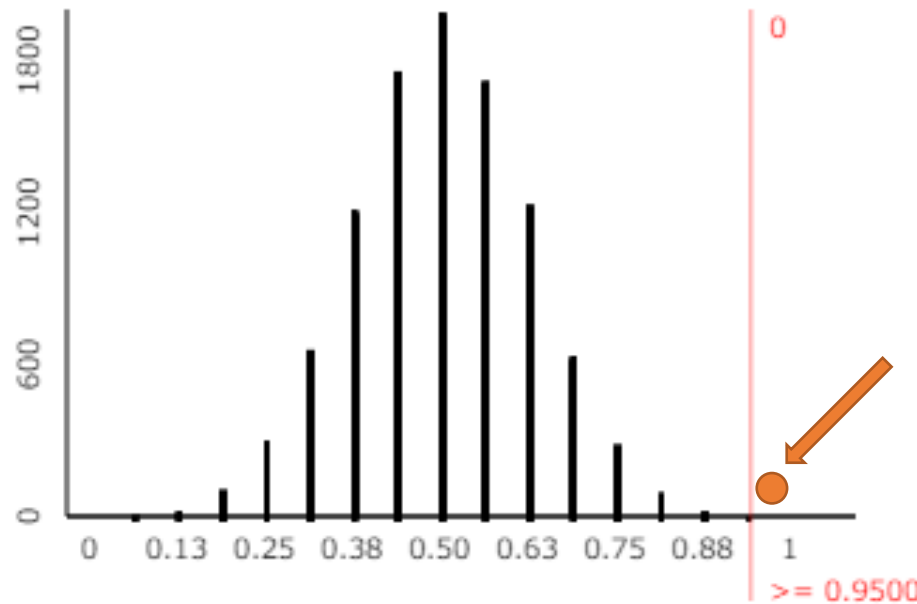
Jimmy A Doi,
Active Learning Lectures for Statistical
Understanding

これなら全部わかる

統計的【① 仮説検定】

主張（赤ちゃんには好みがある） = ② 対立仮説

（赤ちゃんには好みがない） = ③ 帰無仮説



④ 有意水準

⑤ 棄却

⑥ 採択

例えば検定の利用場面

疑問 運動部の人ってみんな体育の時間が好きだ

仮説 運動部に入ると、みんな好きになる？

結果

- ①運動部所属で、運動が好き
- ②運動部所属で、運動が嫌い
- ③運動部以外で、運動が好き
- ④運動部以外で、運動が嫌い

所属部活\好み	好き(14)	嫌い(16)
運動部(15)	8	7
運動部以外(15)	6	9

(期待度数表)

所属部活\好み	好き(14)	嫌い(16)
運動部(15)	7	8
運動部以外(15)	7	8

$$\begin{array}{ccccccc} \text{クラスの数} & \times & \text{運動部の割合} & \times & \text{好きの割合} \\ 30 & \times & \frac{1}{2} & \times & \frac{14}{30} \end{array}$$

期待度数表と比べて、
この表のレア度判定をする

ただし、有意だったとしても...？

2x2に限ってやるのも面白い

適度な「制約」が、「探究」しやすい？

(1)「男女」×「TwitterとInstagramの使用率」

Twitterは主に文章、Instagramは主に写真。情報収集や自分の考えを広めたい男性はTwitterに、暇潰しや様々な人と繋がりたい女性はInstagramに流れる？

(2)「男女」×「Instagramの使用率」

Instagramを使うことは女子力の象徴である。Instagramを使用することで、女子力がアップする。

(3)「妹弟の存在」×「子ども好き」

妹弟がいると、子どもに慣れることができるので子どもを好きになる！

(4)「運動部／運動部以外」×「体育の好み」

体育が好きだから、運動部に入部する！

(5)「男女」×「メイク／すっぴんの好み」

男子には関係が薄く近寄りがたいイメージをメイクは生じさせるが、女子にとってはかわいく生まれ変わりたいという事実で共感できる！

「平均の差の検定」もある

参考

- <https://saikolodsm.com/t-test-and-analysis-of-variance/>

検定までやると、それはそれで生徒の探究も広がる

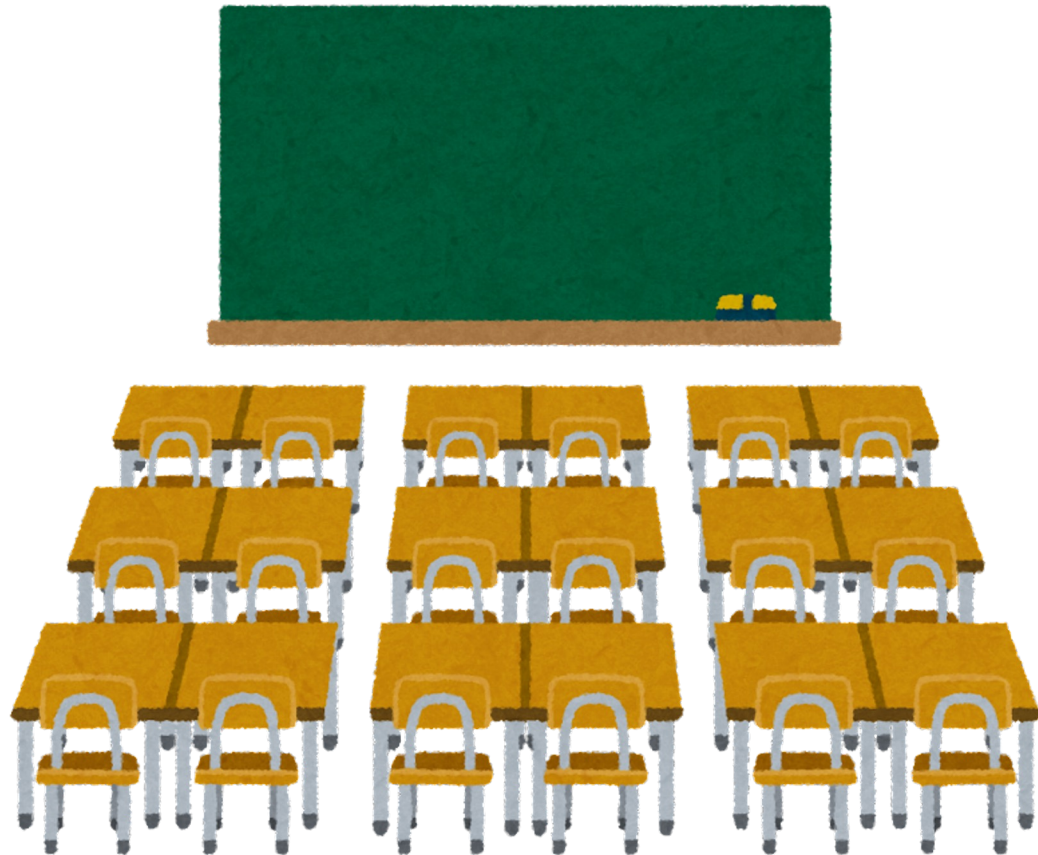
情報リテラシーは情報をむやみに疑うことではないと実感した。

統計を学んだ後になるといろいろなことに疑問がでてくる。
例えば私は歌が下手だが、もし小さい頃からピアノを習っていたら歌が上手くなっていたのかなと思って、ピアノを習っていた人とそうでない人でカラオケボックスで国歌を歌って採点結果を比べてみたら違いがあるのではないかと思った。そこで実際に検定して有意になれば、私は自分の子供には小さい頃からピアノを習わせたいと思う。

③多変量解析の体験

例えば生徒がこんなテーマに取り組んだ時

クラスのみんな、
気が合う人は誰だろう？



アンケートをやってみよう！

どれくらい好きか、選んで下さい。

5 (とても好き) / 4 (好き) / 3 (普通)
2 (あまり好きではない) / 1 (好きではない)

Q1. プログラミング	5	4	3	2	1
Q2. 野球	5	4	3	2	1
Q3. 楽器演奏	5	4	3	2	1
	⋮				
	⋮				
	⋮				

回答結果

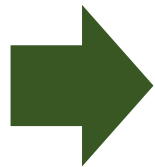
	洋楽	ドラマ	野球	JPOP	プログラミライブ	バスケ	カラオケ	ペンギン	水泳	登山	舞台鑑賞	パン	ダンス	料理	楽器演奏	昼寝	動画鑑賞	犬	猫	
生徒1	3	5	5	2	1	4	1	2	3	3	2	3	4	5	1	4	3	5	5	
生徒2	3	3	3	2	3	1	5	1	2	5	3	5	5	4	2	1	2	2	3	5
生徒3	1	5	4	3	2	3	4	4	1	4	2	4	5	3	3	3	4	5	4	1
生徒4	4	3	5	2	1	3	5	4	1	2	2	2	4	5	3	2	4	4	1	2
生徒5	2	4	2	3	4	1	3	4	3	4	5	4	3	1	1	2	5	2	3	2
生徒6	3	5	1	2	5	2	5	5	2	1	3	5	2	2	2	5	3	3	4	4
生徒7	4	2	1	1	2	5	5	4	3	3	5	2	1	5	1	3	2	1	2	3
生徒8	2	3	3	3	1	3	1	3	4	3	4	5	5	5	4	1	4	1	4	3
生徒9	3	2	4	1	2	2	5	2	4	2	3	2	4	4	5	4	2	4	1	4
生徒10	4	5	4	2	1	2	4	3	3	4	4	4	3	5	5	4	3	4	5	1
生徒11	5	5	1	5	2	3	1	5	3	2	2	3	5	5	3	5	5	2	5	5
生徒12	5	1	1	5	5	4	1	5	4	2	1	4	3	5	1	5	4	4	2	1
生徒13	5	1	1	3	5	2	1	5	2	4	3	1	4	2	5	4	1	3	4	2
生徒14	4	2	4	1	3	1	2	3	4	2	3	5	3	3	5	3	2	1	1	1
生徒15	2	2	2	1	2	3	4	4	4	5	4	2	1	2	4	2	4	5	2	3
生徒16	5	1	4	4	2	3	4	4	3	4	3	1	4	1	4	3	3	2	5	2
生徒17	5	2	4	1	2	1	3	2	4	3	3	2	1	3	4	4	4	4	4	5
生徒18	2	5	2	3	1	3	5	5	1	1	3	3	2	5	3	4	2	3	4	2
生徒19	1	1	3	3	1	5	5	4	2	4	2	1	2	4	4	3	5	5	4	1
生徒20	1	2	2	1	5	2	1	4	2	3	2	1	5	5	5	3	1	3	2	3
生徒21	3	2	4	4	2	1	1	4	2	5	2	1	1	3	3	2	1	5	3	1
生徒22	2	3	5	4	2	2	2	2	2	5	4	5	2	2	4	1	2	4	4	4
生徒23	5	2	2	1	1	1	5	4	5	2	2	1	3	1	5	5	1	3	4	5
生徒24	4	5	2	5	1	4	3	5	4	3	2	3	1	5	5	5	2	5	4	3
生徒25	5	1	4	4	1	1	1	2	1	5	4	5	3	4	5	2	2	5	3	2
生徒26	5	5	2	3	1	2	5	5	1	3	4	3	2	3	2	2	5	2	4	2
生徒27	3	2	4	5	2	4	1	4	1	3	1	4	4	5	2	4	5	5	2	3
生徒28	5	3	2	5	4	2	3	3	2	1	3	5	1	2	5	4	3	5	3	4

どうやって似ている人を探す？

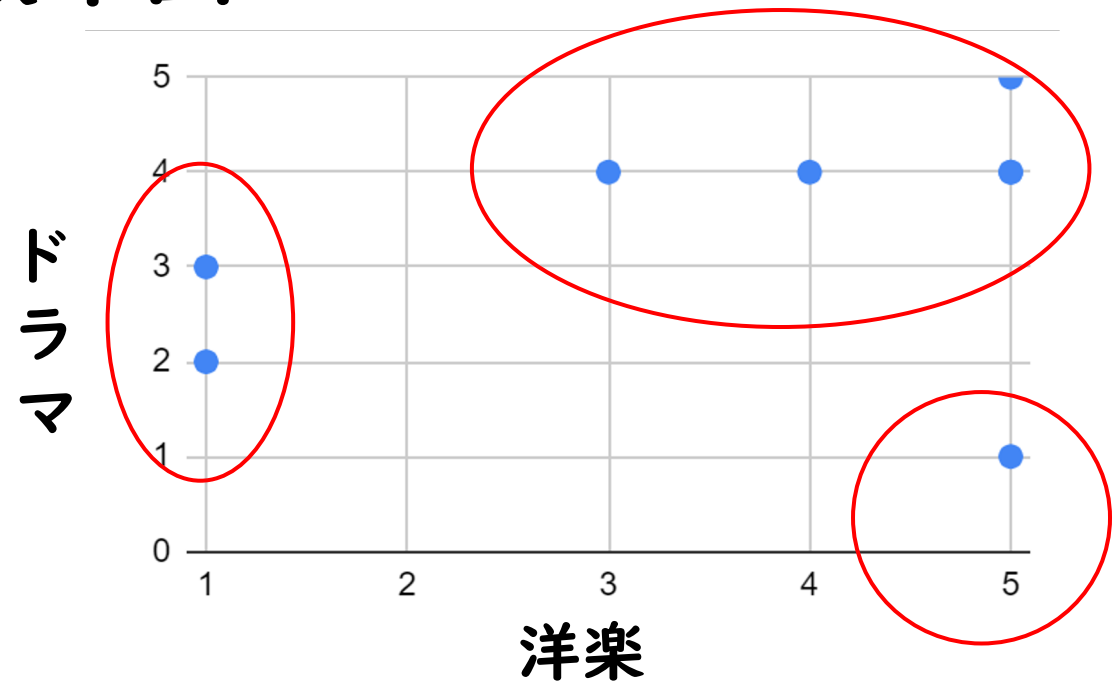
2変数だったら…(情報1でやったこと)

データ

	洋楽	ドラマ
生徒1	3	5
生徒2	3	3
生徒3	1	5
生徒4	4	3
生徒5	2	4
生徒6	3	5
生徒7	4	2
生徒8	2	3
生徒9	3	2
生徒10	4	5
生徒11	5	5
生徒12	5	1



散布図



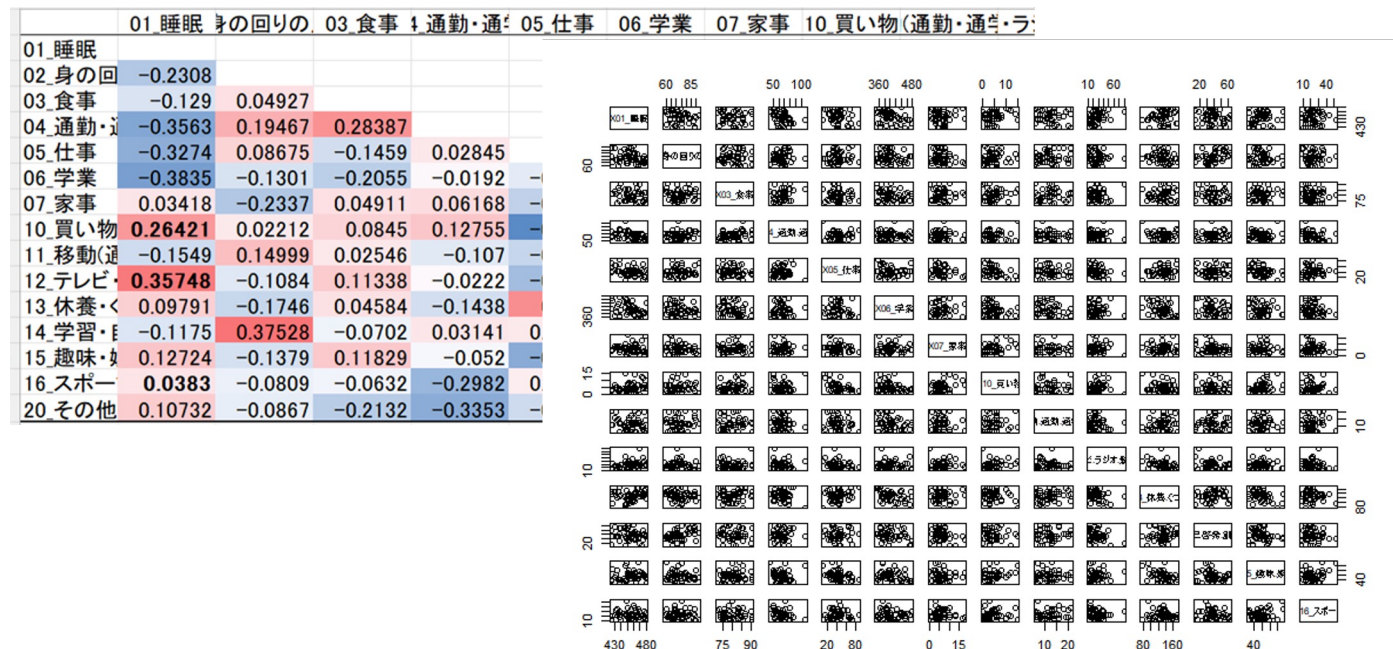
データがもっとたくさんあったら…？

データ

	01_睡眠	02_身の回りの用事	03_食事	04_通勤・通学	05_仕事	06_学業	07_家事	10_買い物	11_移動・通学を除く	12_テレビ・ラジオ・新聞・雑誌	13_休養・くつろぎ	14_学習・自己啓発・娯楽(学業以外)	15_趣味・娯楽	16_スポーツ
01_北海道	467	72	74	62	41	355	5	3	19	25	166	42	63	18
02_青森県	489	71	90	53	34	368	0	0	13	79	128	19	74	30
03_岩手県	450	89	76	53	47	453	5	3	8	31	105	42	59	14
04_宮城県	446	75	80	72	44	373	3	3	7	16	93	29	110	53
05_秋田県	467	61	84	45	29	373	4	6	13	19	136	32	114	34
06_山形県	477	74	83	52	43	377	4	5	6	40	135	38	58	29
07_福島県	436	60	87	64	41	438	7	2	15	23	146	36	40	25
08_茨城県	456	88	76	71	31	387	6	6	18	24	133	41	48	35
09_栃木県	472	66	82	69	46	394	13	12	9	24	131	13	72	20
10_群馬県	457	74	78	69	41	398	7	7	12	31	125	36	67	22
11_埼玉県	439	83	91	80	31	401	5	3	12	28	134	55	41	13
12_千葉県	437	71	85	85	70	387	8	3	12	15	170	54	33	17
13_東京都	451	84	86	80	33	387	5	7	11	16	124	55	56	36
14_神奈川県	452	89	86	77	32	399	2	8	14	39	108	23	78	9
15_新潟県	475	71	74	50	13	361	7	11	6	49	143	49	62	21
16_富山県	475	68	75	66	88	376	3	4	10	35	125	28	64	15
17_石川県	460	69	78	51	48	395	3	1	16	16	135	31	79	43
18_福井県	445	87	75	56	36	450	3	4	10	12	126	54	45	25
19_山梨県	453	75	78	62	28	429	5	1	16	30	94	60	66	14
20_長野県	463	62	91	72	43	402	10	1	6	26	150	30	53	17
21_岐阜県	458	73	78	69	38	418	13	6	9	45	101	36	57	23
22_静岡県	452	82	81	62	48	390	8	3	11	21	136	41	54	23
23_愛知県	459	93	79	71	72	374	4	1	12	25	123	44	51	13
24_三重県	437	68	75	81	87	358	3	2	6	21	171	42	41	23
25_滋賀県	462	81	83	60	47	398	5	4	6	12	144	46	47	27
26_京都府	478	65	83	53	23	360	17	10	14	27	159	34	53	15
27_大阪府	430	88	84	70	65	366	8	6	16	23	147	56	51	22
28_兵庫県	465	79	87	69	45	351	3	6	12	33	161	40	48	19
29_奈良県	446	72	86	107	10	416	7	15	9	41	75	36	100	13
30_和歌山県	467	58	81	71	31	403	7	6	14	37	139	31	45	23
31_鳥取県	464	69	72	55	55	408	9	4	13	37	138	38	27	31
32_島根県	449	68	75	66	47	440	1	1	8	14	152	51	37	19
33_岡山県	424	89	82	77	40	405	4	2	21	14	135	49	40	11
34_広島県	450	64	80	67	27	433	18	6	7	20	132	28	71	17
35_山口県	432	70	81	67	63	384	7	5	20	25	133	34	63	32
36_徳島県	451	76	77	52	65	401	3	3	8	31	111	61	64	18



相関行列や散布図



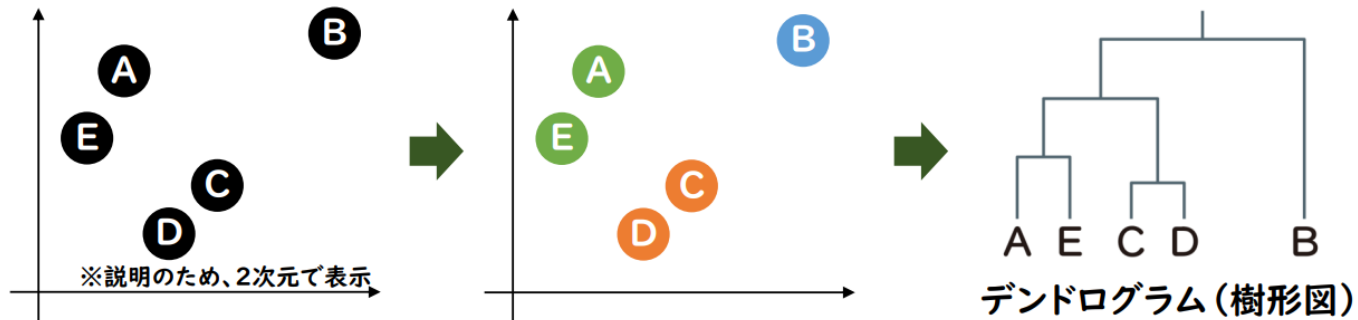
結局全体的には何を言えいいの？

→こんな時に役立つ1つの方法が、**クラスタリング**

高校数学(中学数学)の範囲内でイメージ可能

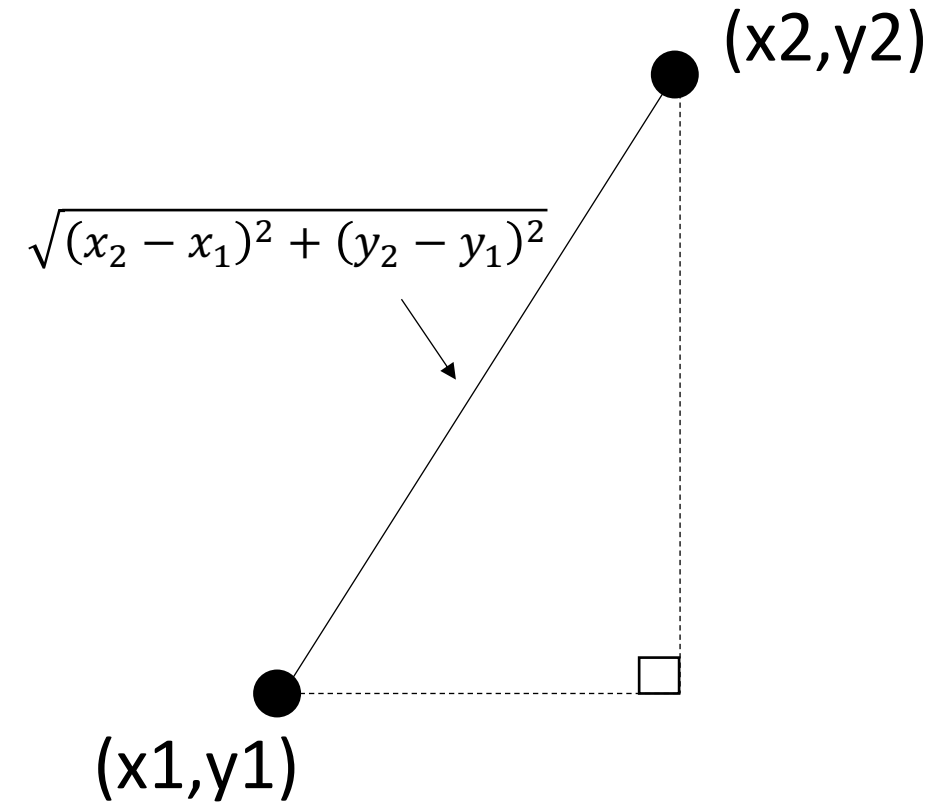
クラスタリング

データ間の距離を計算して、近いもの同士を結びつけて行く。



3次元以上の
データにも対応!

	洋楽	ドラマ	野球	JPOP	プログラミライブ	バスケ	カラオケ	ペンギン	水泳	登山	舞台鑑賞	パン	ダンス
生徒1	3	5	5	2	1	4	1	2	3	3	2	3	
生徒2	3	3	3	2	3	1	5	1	2	5	3	5	5
生徒3	1	5	4	3	2	3	4	4	1	4	2	4	5
生徒4	4	3	5	2	1	3	5	4	1	2	2	2	4
生徒5	2	4	2	3	4	1	3	4	3	4	5	4	3
生徒6	3	5	1	2	5	2	5	5	2	1	3	5	2
生徒7	4	2	1	1	2	5	5	4	3	3	5	2	1
生徒8	2	3	3	3	1	3	1	3	4	3	4	5	5
生徒9	3	2	4	1	2	2	5	2	4	2	3	2	4
生徒10	4	5	4	2	1	2	4	3	3	4	4	4	3
生徒11	5	5	1	5	2	3	1	5	3	2	2	3	5
生徒12	5	1	1	5	5	4	1	5	4	2	1	4	3
生徒13	5	1	1	3	5	2	1	5	2	4	3	1	4
生徒14	4	2	4	1	3	1	2	3	4	2	3	5	3
生徒15	2	2	2	1	2	3	4	4	4	5	4	2	1



三平方の定理!

今回使用する環境

オープンソースの統計解析向けプログラミング言語



コマンドの解説(一時停止して見て下さい)

```
syumi <- read.csv("syumi.csv",h=T,row.names=1)
```

変数「syumi」に「syumi.csv」ファイルを読み込む。

h=T ヘッダが存在する / row.names=1 1列目のデータを行の名前にする。

```
syumi.d <- dist(syumi)
```

dist関数を利用して、データ間の距離を計算する(distance=距離)。

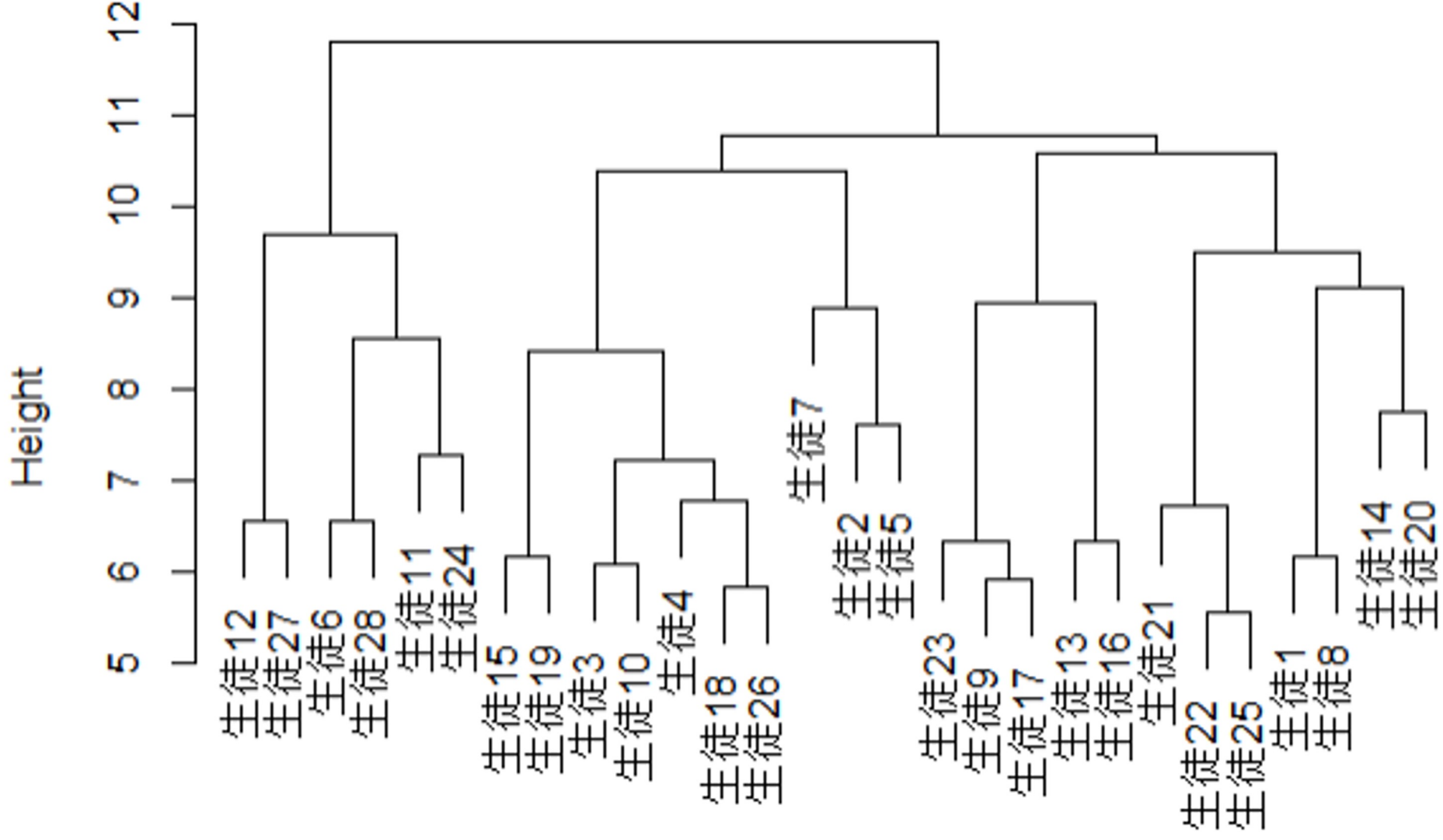
```
syumi.c <- hclust(syumi.d)
```

hclust関数を利用して、**クラスタリング**を行う。

ヘッダ

	洋楽	ドラマ	野球	JPOP	プログラミライブ	
生徒1	3	5	5	2	1	4
生徒2	3	3	3	2	3	1
生徒3	1	5	4	3	2	3
生徒4	4	3	5	2	1	3
生徒5	2	4	2	3	4	1
生徒6	3	5	1	2	5	2
生徒7	4	2	1	1	2	5
生徒8	2	3	3	3	1	3
生徒9	3	2	4	1	2	2

クラスタリングの結果

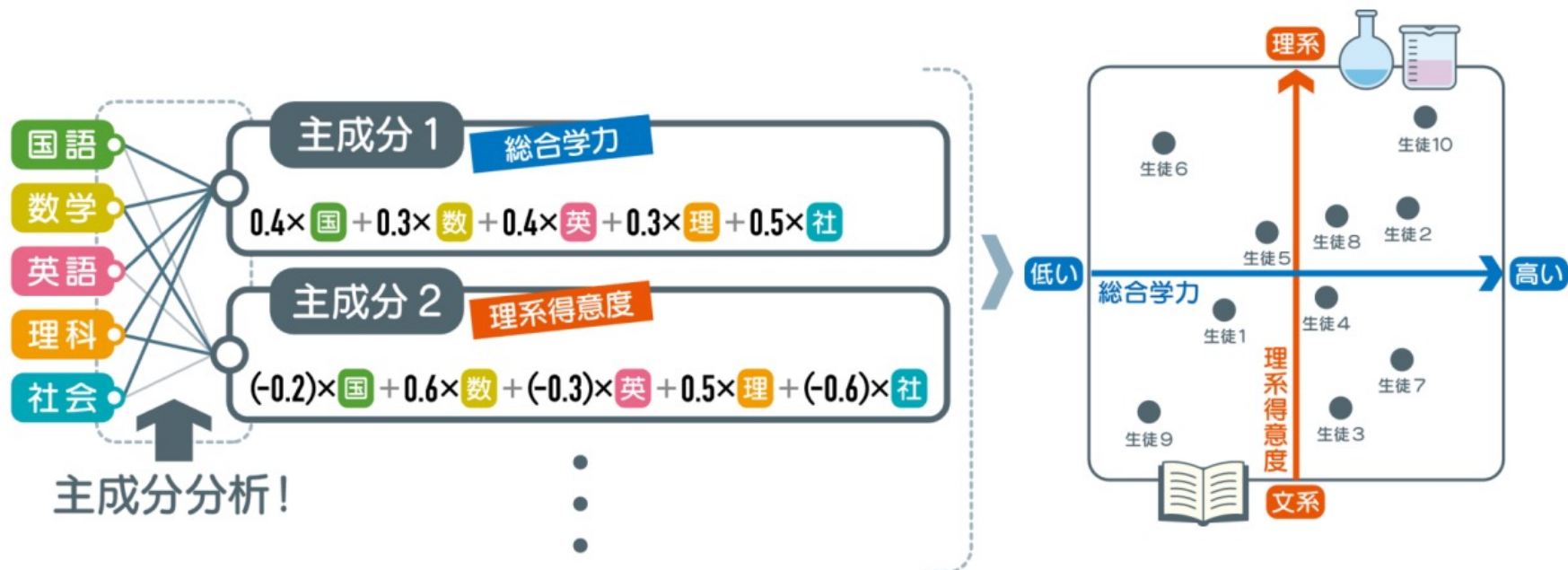


その他に多次元データを圧縮表現する方法→主成分分析

<https://www.youtube.com/watch?v=rKwjOnzXw8U>

主成分分析

多次元のデータを低次元の主成分に圧縮（要約）することで、データの構造を可視化できる。

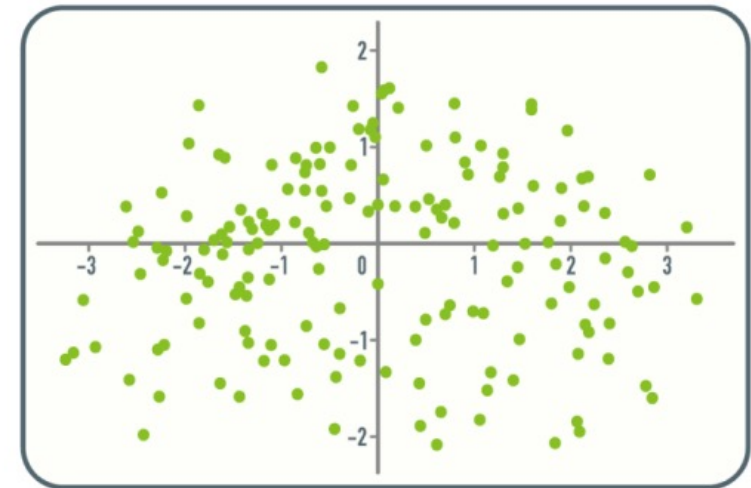
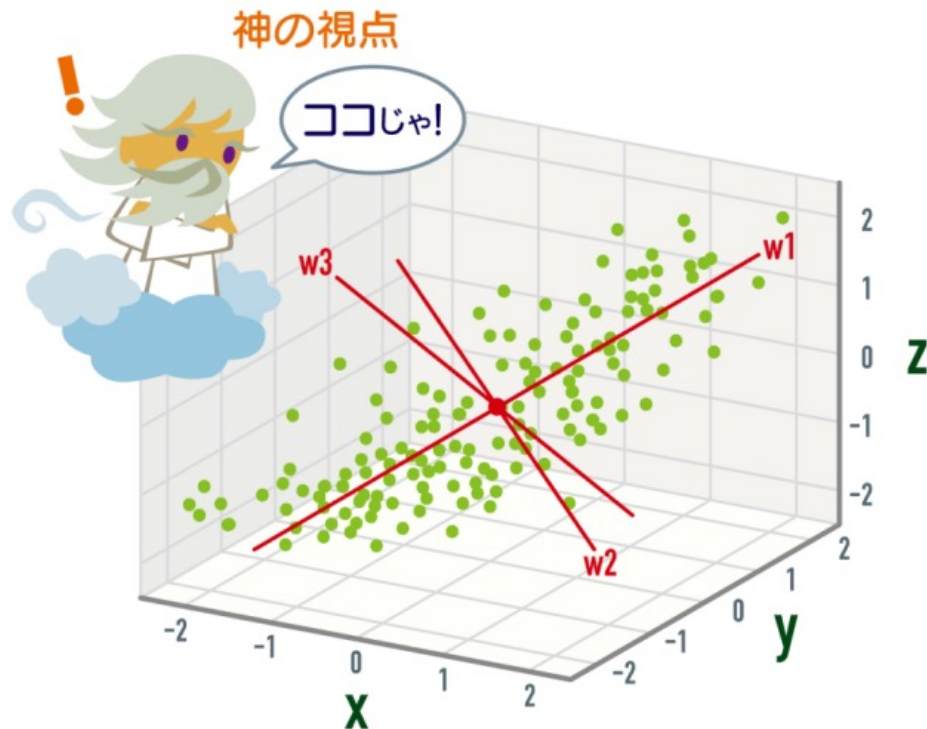


主成分分析の仕組み

<https://www.youtube.com/watch?v=rKwjOnzXw8U>

圧縮（要約）のイメージ

散らばりを最大化するように、空間上で軸を決める。
(一番散らばって見えるところで、見下ろすイメージ)



文科省「情報Ⅱ」解説動画より

文科省動画でやった例



独立行政法人
統計センター



<https://www.nstac.go.jp/use/literacy/ssdse/>

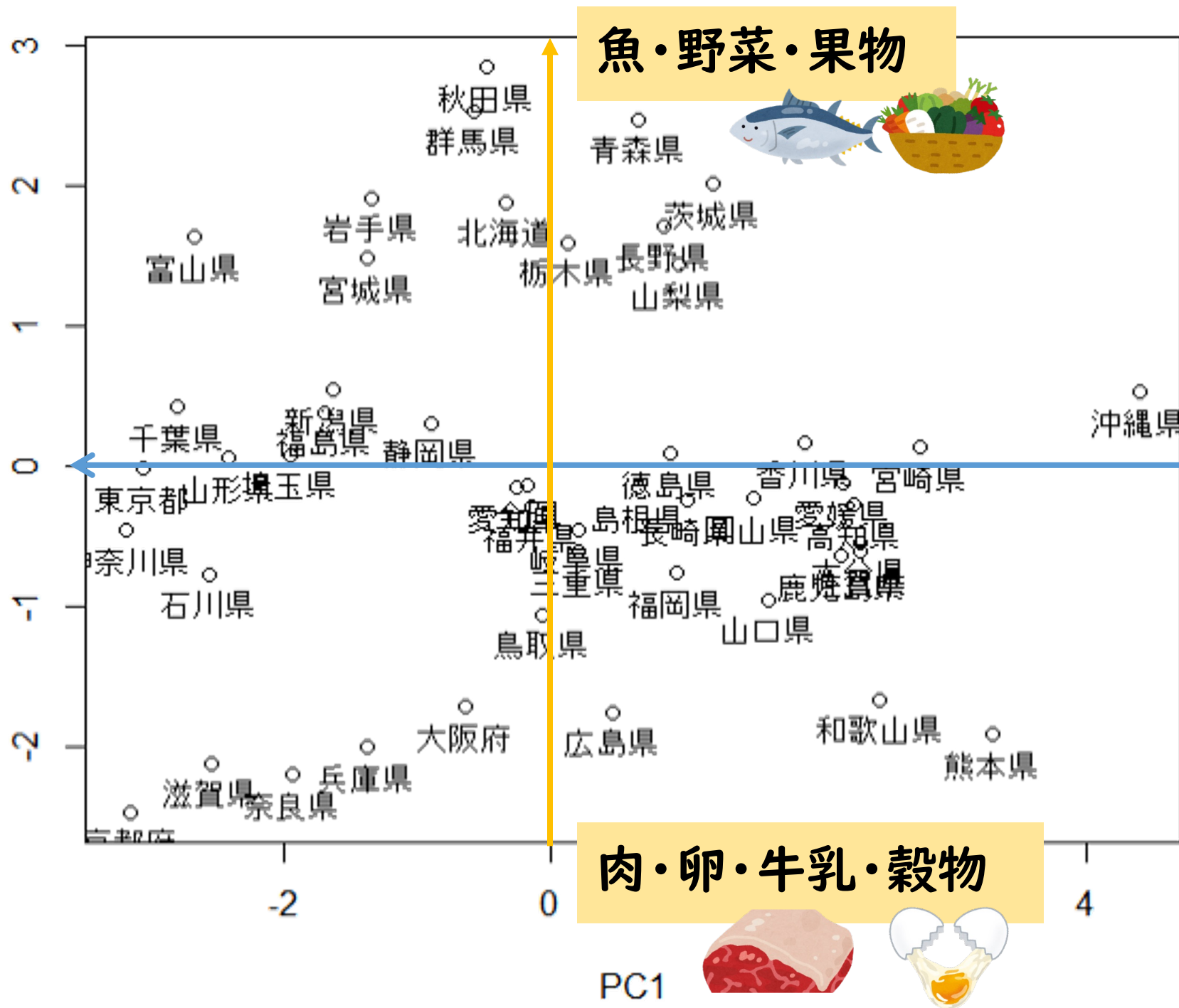
	01 穀類	02 魚介類	03 肉類	牛乳	卵	05 野菜	06 果物	08 菓子類
全国	79597	75171	97501	15285	10250	108315	40424	89367
北海道	80069	83434	91587	12873	9054	108284	43848	89616
青森県	74692	87151	89759	13751	7876	112569	38445	82689
岩手県	79493	77147	81637	16556	8921	116270	44146	94418
宮城県	76748	80756	86259	14886	10044	119365	44600	96002
秋田県	70765	83200	84768	13326	9326	120467	47823	88991
山形県	81774	75220	104651	16811	10001	116856	49520	95511
福島県	76382	76925	85656	15018	12258	119124	49191	94925
茨城県	71011	67538	79669	13656	9485	103200	42808	92725
栃木県	76722	70647	85784	14338	9254	110330	42586	93030
群馬県	82055	72087	77366	14965	8706	109748	46874	90262
埼玉県	82383	73742	98466	16105	9845	118512	42548	101101
千葉県	82395	80432	93747	17247	10308	126613	46139	95294
東京都	81898	82370	106706	16524	10137	129296	46646	96423
神奈川県	86334	81864	105787	16034	10637	131479	45025	95843
新潟県	86628	75021	86221	15470	10445	126203	42360	90363
富山県	86991	88471	91485	15420	8888	119643	44898	95398
石川県	87677	76898	103972	16577	9948	111969	40422	105512

SSDSE-家計消費 (SSDSE-C)

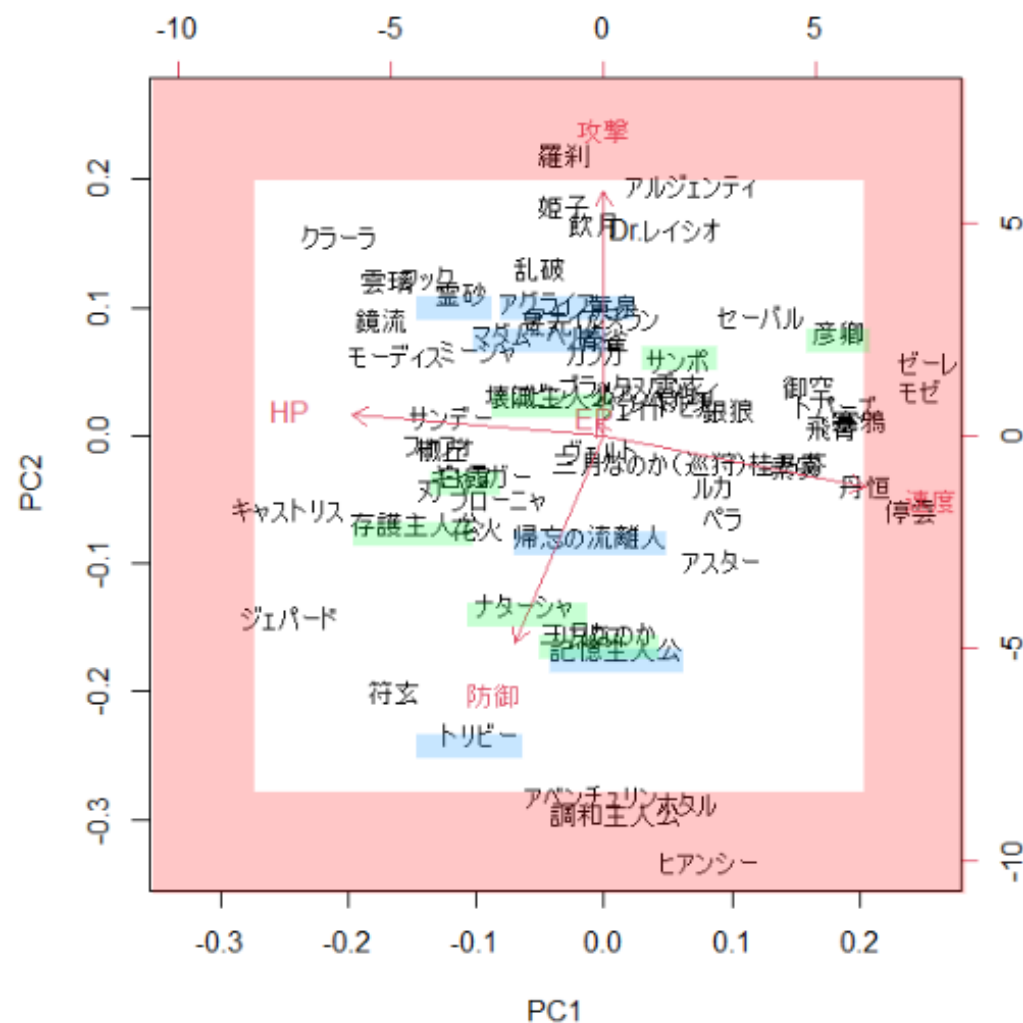
- 都道府県庁所在市別の家計消費データを集めたデータセットのうち、一部。
- 都道府県庁所在市別 (→都道府県に読み替え)
- 二人以上の世帯の1世帯当たり (≡人口は関係なし)
- 品目別 (食料の全品目) 年間支出金額を収録。

課題「都道府県ごとに、各食品の消費には
どんな傾向があるだろう？」





生徒の作例



赤い範囲の中は比較的评价が高いとされるキャラ

评价が低いキャラ

评价が高いキャラ

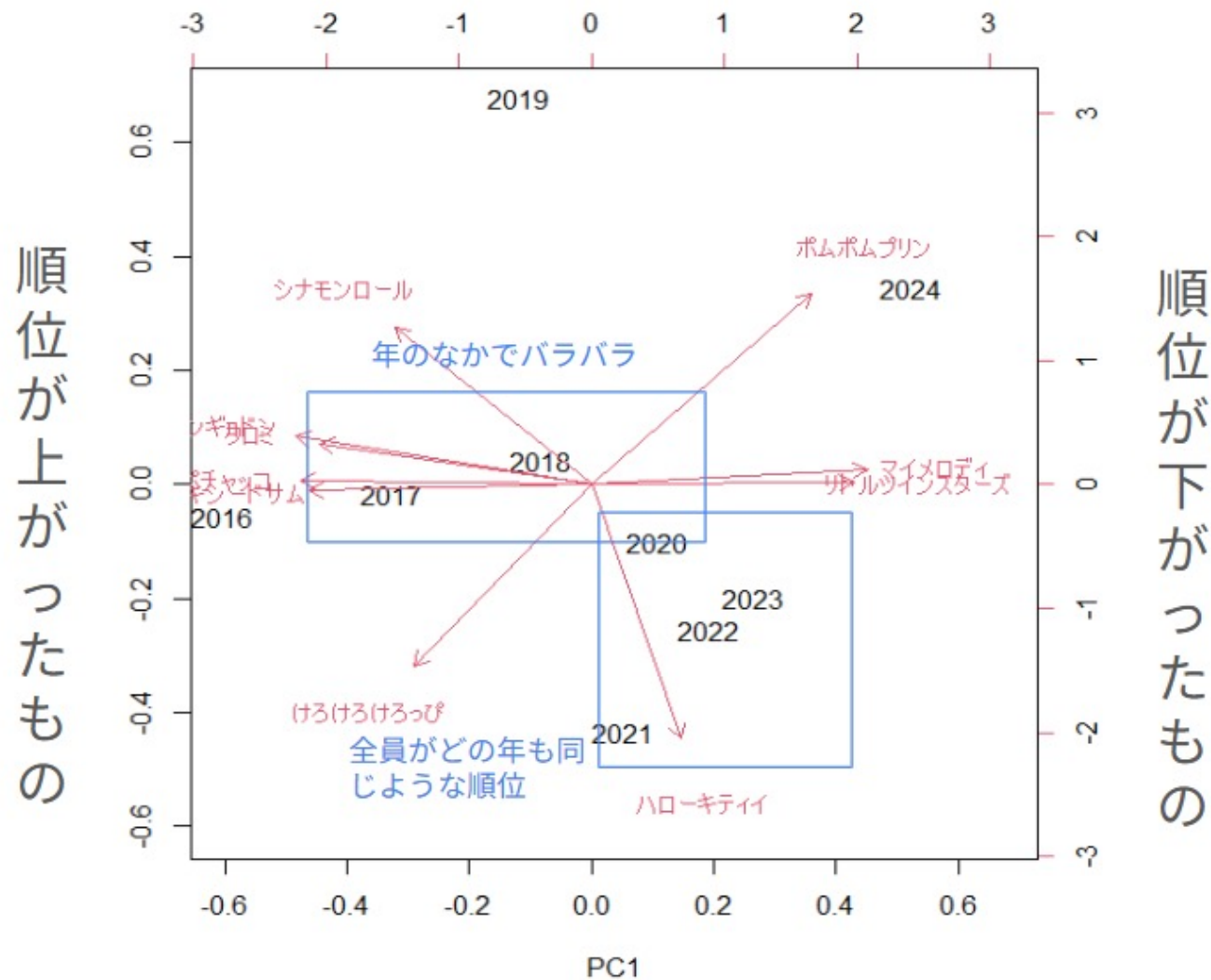
活用したデータ
崩坏スターレイルキャラ性能

使用した変数
攻撃 HP 防御 速度

実習を通じて得られた知見

- 中心から外に行くにつれて、必ずプレイヤー評価が高い訳では無い
- 全体的にステータスがバラけていた

生徒の作例②



「情報Ⅱ」という科目では

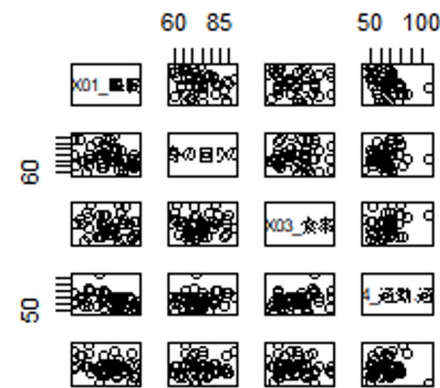
多次元のデータ

	01. 睡眠	02. 身の回りの	03. 食事	04. 通勤・通学	05. 仕事	06. 学業	07. 家事	08. 趣味・娯楽	09. 運動・スポーツ	10. 健康・美容	11. 経済・金融	12. 社会・環境	13. 自然・地理	14. 文化・芸術	15. その他
01. 睡眠	467	72	74	62	41	355	5	9	19	25	166	42	63	18	
02. 身の回りの	469	71	80	53	34	388	9	9	15	79	129	19	14	30	
03. 食事	459	69	76	53	47	453	5	3	9	31	105	42	59	14	
04. 通勤・通学	448	65	60	72	44	373	3	3	6	50	50	33	115	53	
05. 仕事	467	61	84	45	29	373	4	6	10	19	136	32	114	34	
06. 学業	477	63	69	42	40	377	4	9	6	40	135	38	58	29	
07. 家事	438	60	87	64	41	438	7	2	15	29	148	38	40	25	
08. 趣味・娯楽	466	66	70	71	31	387	8	6	10	24	133	41	48	20	
09. 運動・スポーツ	472	68	62	69	46	384	13	12	5	24	131	13	72	20	
10. 健康・美容	457	74	78	68	41	388	7	7	12	31	125	38	67	22	
11. 経済・金融	459	63	91	60	31	401	5	3	12	28	154	35	41	13	
12. 社会・環境	437	71	85	65	70	387	6	3	12	15	170	54	33	17	
13. 自然・地理	451	64	66	60	33	387	5	7	11	16	124	55	56	36	
14. 文化・芸術	452	69	86	77	32	388	2	8	14	39	108	23	78	9	
15. その他	475	71	74	56	12	391	7	11	9	48	143	49	62	21	
16. 睡眠	475	66	75	66	68	376	3	4	10	35	125	29	64	15	
17. 身の回りの	469	69	78	51	48	395	3	1	16	16	135	31	75	45	
18. 食事	445	67	75	56	38	450	3	4	10	12	129	54	45	25	
19. 通勤・通学	453	75	78	62	28	429	5	1	16	30	94	68	68	14	
20. 仕事	463	62	91	72	43	402	10	1	9	39	150	39	53	17	
21. 学業	459	73	78	69	38	418	13	6	9	45	101	36	57	23	
22. 家事	452	62	91	62	49	390	9	3	11	21	138	47	54	23	
23. 趣味・娯楽	459	63	79	71	72	374	4	1	12	25	123	44	51	13	
24. 運動・スポーツ	471	66	67	67	59	390	3	6	11	17	144	42	41	23	
25. 健康・美容	462	61	83	60	47	388	5	4	6	12	144	46	47	27	
26. 経済・金融	476	65	83	53	23	380	17	10	14	27	159	34	53	15	
27. 社会・環境	450	68	64	70	65	386	6	6	15	23	147	56	51	22	
28. 自然・地理	465	79	87	69	45	351	3	6	12	33	161	40	48	18	
29. 文化・芸術	446	72	96	107	10	418	7	15	3	41	75	36	100	13	
30. その他	467	58	91	71	31	452	7	6	14	37	139	31	45	23	
31. 睡眠	464	69	72	55	55	400	9	4	13	37	138	38	27	31	
32. 身の回りの	449	68	75	66	47	440	1	1	9	14	152	51	37	19	
33. 食事	451	66	82	77	40	405	4	2	21	14	135	48	48	11	
34. 通勤・通学	450	64	80	67	27	433	18	6	7	20	152	39	71	17	
35. 仕事	432	70	81	67	63	384	7	5	20	25	133	34	63	32	
36. 学業	451	78	77	72	65	401	3	3	31	111	61	64	38	18	

情報Ⅰ

2つずつで
頑張ってみる

	01. 睡眠	02. 身の回りの	03. 食事	04. 通勤・通学	05. 仕事	06. 学業
01. 睡眠						
02. 身の回りの	-0.2308					
03. 食事	-0.129	0.04927				
04. 通勤・通学	-0.3563	0.19467	0.28387			
05. 仕事	-0.3274	0.08675	-0.1459	0.02845		
06. 学業	-0.3835	-0.1301	-0.2055	-0.0192	-0.0992	
07. 家事	0.03418	-0.2337	0.04911	0.06168	-0.1948	0.05732
08. 趣味・娯楽	0.03418	-0.2337	0.04911	0.06168	-0.1948	0.05732
09. 運動・スポーツ	0.26421	0.02212	0.0845	0.12755	-0.4664	-0.239
10. 健康・美容	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163
11. 移動・通学	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163
12. テレビ	0.35748	-0.1084	0.11338	-0.0222	-0.2739	-0.0922
13. 休養・くつろぎ	0.09791	-0.1746	0.04584	-0.1438	0.2955	-0.4069
14. 学習・読書	-0.1175	0.37528	-0.0702	0.03141	0.02206	-0.2363
15. 趣味・娯楽	0.12724	-0.1379	0.11829	-0.052	-0.3413	-0.2828
16. スポーツ	0.0383	-0.0809	-0.0632	-0.2982	0.01111	-0.1535
20. その他	0.10732	-0.0867	-0.2132	-0.3353	-0.1576	0.01509



相関行列

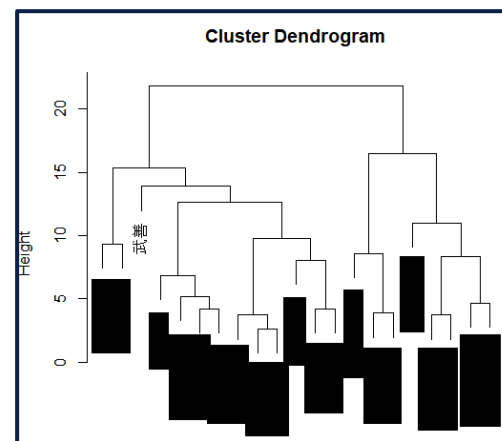
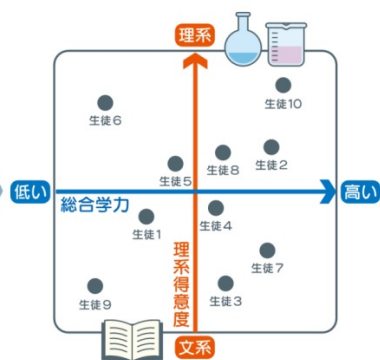
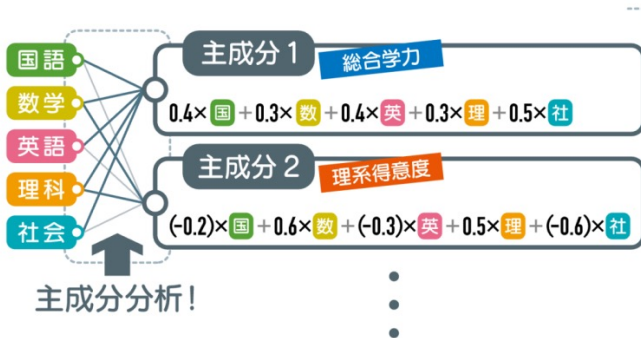
散布図行列

情報Ⅱ

そのまま見る!

主成分分析で、次元を圧縮しよう

クラスタリングで、グループ化しよう



etc...

情報Ⅰ／情報Ⅱ解説動画 プロジェクト

文科省「情報Ⅰ」「情報Ⅱ」解説動画 <https://www.nttls-edu.jp/joho/>

- ・ データサイエンス領域で情報Ⅰ、情報Ⅱともに関わった
- ・ 「これぞ情報Ⅰ」「これぞ情報Ⅱ」と言える内容にはなったと思われる

情報Ⅰ



情報Ⅱ



6.本日のまとめ

お疲れ様でした！

本日のまとめ

① 探究には「型」がある

問い → 計画 → データ → 分析 → 結論
PPDACを回し、次の問いにつなげる。

② 大切なのは「分析」だけではない

問いとデータは合っているか。
その結果から、本当にそこまで言えるか。
データを集める前と、分析した後こそ考える。

③ 生徒の探究に「次の一手」を返せるようにする

「その設問だと、欲しいデータにならないかも」
「それは相関であって、因果とはまだ言えない」
「このデータなら、こんな分析も試せそう」
「次は、こんなデータを集めてみたら？」

どのフェーズでもAIが使える。
教員側のファシリテートが
試される。

まずは、自分で回してみる。
その経験が、
生徒の探究を支える「勘所」になる。

**試行錯誤するワクワクを
大切にしていきましょう！**