



教員向け指導力向上事業のための学習会 ～データの種類に応じた分析・教育実践事例～



早稲田大学 創造理工学部 経営システム工学科
蓮池 隆

2023年1月19日(木)

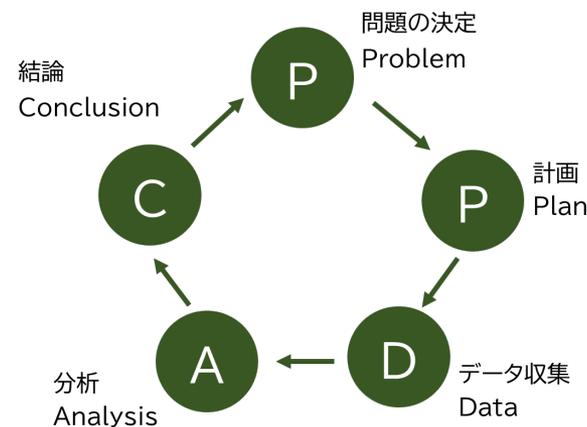
本講演に関して

- 本講演は以下の2つの授業解説動画のフォローアップ、補足解説，発展解説を目的としています
- 情報通信ネットワークとデータの活用(4)
「アンケートで身近な問題を解決しよう！」
<https://www.youtube.com/watch?v=VFzCUPp30jg&t=41s>
- 情報通信ネットワークとデータの活用(5)
「オープンデータの分析」
(近日公開予定)
- 本日は**データ分析のPPDACサイクル「データをどう収集し、分析し、サイクルを回すのか？」**の流れを見ていきます

(前回の一部再掲)

『情報I』におけるデータサイエンス

- 学習指導要領：(4)情報通信ネットワークとデータの活用
 - 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
 - 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 → **データ分析**
- データ分析に関わる一連の流れを，実際にデータを操作・分析しながら理解を深め，身に付けることが目的



(前回の再掲)

『情報I』におけるデータサイエンス

	「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, <u>交絡因子</u> 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, <u>単回帰分 析</u> , 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		<u>量的データ</u> 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		<u>質的データ</u> 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 <u>整理されていない データ</u> も扱い, <u>外れ値</u> , <u>欠損値</u> な どの処理も学ぶ
尺度	—		名義, 順序, 間隔, 比例など <u>尺度 水準の違い</u> を扱う
データベース	「情報の科学」のみで扱う		<u>情報を収集・蓄積・提供する方法 として全員が学ぶ</u>

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 赤字 = 情報科のみで活用

(前回の再掲) データ分析にトライしよう

- おそらく、前回の学習指導要領からの大きな変更点がこの「**データ分析**」「**データを用いた問題解決**」
 - 「データ分析は難しそう…」
 - 「今までやったことないのにできるか不安…」
 - 「データ分析って何をどうすればよいの？」
- データ分析は怖いものではけっしてありません！
難しい部分もあるかもしれませんが、**失敗を恐れずに
トライしてみることが重要**



いただいた質問内容から…①

- **実習に用いるデータは、どこから探せばよいか？**
 - **オープンデータ**の利用(公的な統計調査データやスポーツ系のデータは比較的取りやすい)
 - オープンデータだけでもいろいろな解析が可能
- 利用可能なオープンデータの例
 - 公的機関(例：総務省統計局, RESAS, 市町村のWebページ)
 - データカタログサイト
https://www.data.go.jp/#!#page:1&frame_id:55
 - (登録すれば) NII 情報学研究データリポジトリ
 - (海外でもよければ) UCIオープンデータセット

例：総務省統計局

統計局ホームページ/労働力調査・家計調査

https://www.stat.go.jp/info/kouhou/keijyou.html

Google カスタム検索 検索

→採用情報 →リンク集 →ご意見・お問合せ →サイトマップ →文字サイズ等の変更 →English

Statistics Japan

ホーム 実施中の調査 統計データ よくある質問 統計研究研修 広報・募集 組織紹介

ホーム > インフォメーション > 広報資料 > 統計調査のご案内 > 家計調査・労働力調査・小売物価統計調査のご紹介

変化が見える、暮らしに役立つ

とうけい ちょうさ

統計調査

国が実施する調査です

家計が見える

かけいちょうさ

家計調査

家庭のお金がどのように使われているかを明らかにします。

雇用が見える

ろうどうりょくちょうさ

労働力調査

人々が働いている状況、失業の状況などを明らかにします。

物価が見える

こうりぶつかとうけいちょうさ

小売物価統計調査

モノやサービスの価格の変化などを明らかにします。

例：e-Stat

The screenshot shows the e-Stat website homepage. At the top, there's a navigation bar with the e-Stat logo and the text "統計で見る日本" (View Japan with Statistics). Below this, there are several main sections:

- 統計データを探す (政府統計の調査結果を探します)**: A section for finding statistics, featuring buttons for "すべて" (All), "分野" (Field), and "組織" (Organization). Below these is a search bar with the example "例：国勢調査" and a "検索" (Search) button.
- 統計データを活用する**: A section for utilizing statistics, featuring buttons for "グラフ" (Graph), "時系列表" (Time Series Table), "地図" (Map), and "地域" (Region).
- その他の絞り込み**: A button for further filtering options.
- 利用ガイド**: A button for the user guide.
- 統計データの高度利用**: A section for advanced use, featuring buttons for "マイクロデータの利用" (Use of Microdata) and "開発者向け" (For Developers).
- 統計関連情報**: A section for related information, featuring a button for "統計分類・調査計画等" (Statistics Classification, Survey Plans, etc.).

At the bottom of the page, there are three main sections:

- 新着情報** (New Information): A section for the latest news, with a "一覧表示" (List View) button.
- 公表予定** (Publications Scheduled): A section for upcoming publications, with a "一覧表示" (List View) button.
- ランキング** (Ranking): A section for rankings, with a "検索キーワード" (Search Keyword) dropdown and a "一覧表示" (List View) button. The current ranking is for "国勢調査" (Census) with a value of 3145.

There is also a featured advertisement for a "DataScienceOnlineCourse" titled "誰でも使える統計オープンデータ" (Statistics Open Data Anyone Can Use), which is scheduled for January 10, 2025, and is free of charge.

例：RESAS(地域経済分析システム)

RESAS

☰ マップを選択してください

投票モデル ダッシュボード データ分析支援 サマリー ヘルプ 問合せ

RESAS

Regional Economy Society Analyzing System

地域経済分析システム

地域経済分析システムRESASの
利活用サイト

RESAS Portal

V-RESAS

新型コロナウイルス感染症が
地域経済に与える影響の可視化

RESAS

動画

地方創生★
政策アイデア
コンテスト2022

2022.12.27 Updated

ニュース・イベント

2022.07.12 Updated

関連サービス

2022.07.12 Updated

RESASとは・関連情報

<https://resas.go.jp/#/27/27100>

例 : UCI Machine Learning Repository



[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

Search

Repository Web

Google

[View ALL Data Sets](#)

Check out the [beta version](#) of the new UCI Machine Learning Repository we are currently testing! [Contact us](#) if you have any issues, questions, or concerns. [Click here to try out the new site.](#)



Browse Through: 622 Data Sets

Table View [List View](#)

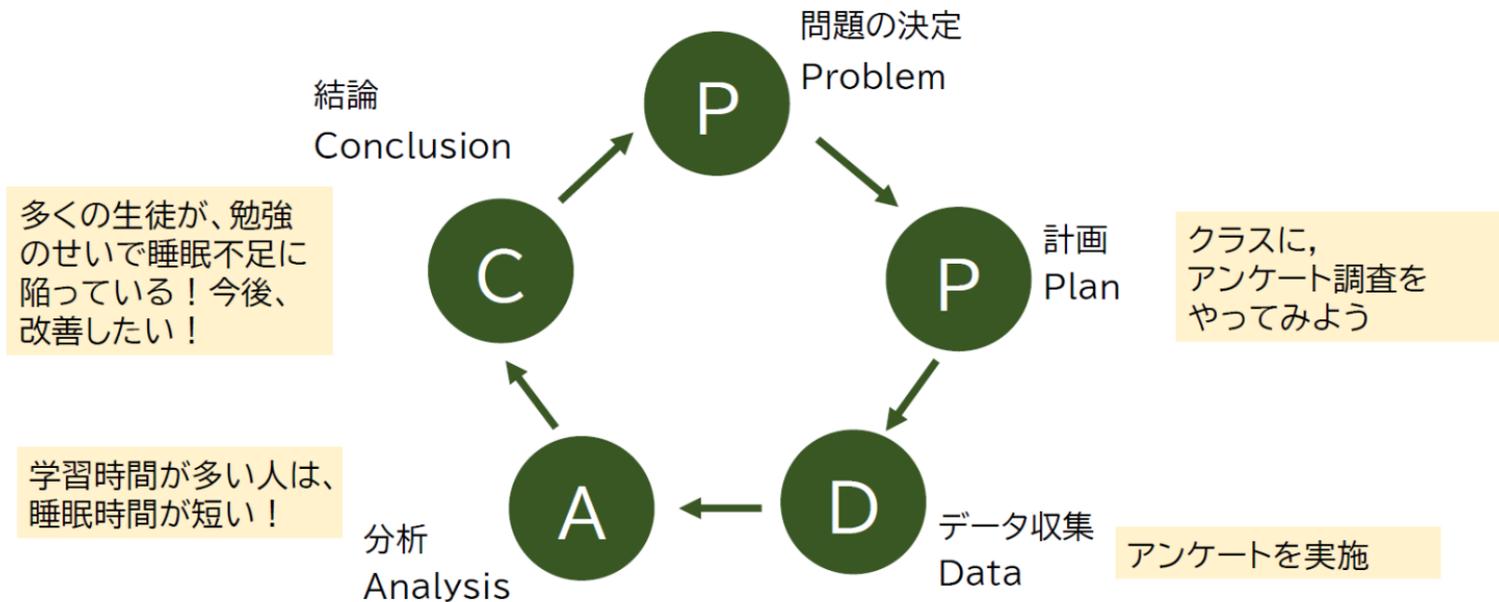
Default Task	Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
Classification (466) Regression (151) Clustering (121) Other (56)	 Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
Attribute Type Categorical (38) Numerical (422) Mixed (55)	 Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
Data Type Multivariate (480) Univariate (30) Sequential (59) Time-Series (126) Text (69) Domain-Theory (23) Other (21)	 Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Area Life Sciences (147) Physical Sciences (57) CS / Engineering (234) Social Sciences (41) Business (45) Game (12) Other (81)	 Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
# Attributes Less than 10 (166) 10 to 100 (279) Greater than 100 (110)	 Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998
# Instances Less than 100 (38) 100 to 1000 (210) Greater than 1000 (339)	 Artificial Characters	Multivariate	Classification	Categorical, Integer, Real	6000	7	1992
Format Type Matrix (439)	 Audiology (Original)	Multivariate	Classification	Categorical	226		1987
	 Audiology (Standardized)	Multivariate	Classification	Categorical	226	69	1992
	 Auto MPG	Multivariate					

<https://archive.ics.uci.edu/ml/datasets.php>

統計データを利用してみよう

- 総務省統計局のデータに関しては、もうすぐ公開されている授業動画(情報通信ネットワークとデータの活用(5)「オープンデータの分析」)でも分析で利用

情報通信ネットワークとデータの活用(4)
「アンケートで身近な問題を解決しよう！」



統計データを利用してみよう

- 授業動画とは異なる視点で、オープンデータから分析
- 総務省統計局の「社会生活基本調査」は多様なデータが収集されているので、分析も行いやすい

今回のテーマ

- 「若者はスマートフォンやPCから情報を得たり、動画を見たりしているから、テレビ(新聞・雑誌)離れが進んでいる？」
 - 休日(日曜日)の動向を「社会生活基本調査」のデータ(65-1)を基に分析

統計データを利用してみよう

- 「若者はスマートフォンやPCから情報を得たり，動画を見たりしているから，テレビ(新聞・雑誌)離れが進んでいる？」
→ 休日(日曜日)の動向を「社会生活基本調査」のデータ(65-1)を基に分析

(結果：差はなさそうだけれど…)

- 「テレビ」だけのデータで分析すると結果が変わるかも？
- そもそもサンプル数が少なすぎる？
- 他に同じような分析をしていないだろうか？(他のデータと組み合わせて分析できないだろうか？)
- スマートフォンやPCの時間が長いと，他の時間が削られるはず → どの時間を削っているのだろうか？

アンケート分析はやはり重要？

- 生徒にとって身近な(学校や住んでいる地域などの)疑問を分析しやすい
- アンケートを作成すること自体がPPDACサイクルを回すことにつながりやすい
 - 疑問解決のためにアンケートを作成する
 - アンケートをとって分析するがあまり良い結果が出ない(想定通りにいかない)
 - 聞いた内容や自由記述欄を見ると、アンケートの不備が発見できる
 - 不備をなくしたアンケートを再実行する
 - …(以下繰り返し)

授業動画を詳細に見ていく

データを問題の発見・解決に
活用する方法

Try!

どんなテーマでアンケート実習をするか、決めてみよう!

(例) (目的) 高校生の生活の実態を明らかにする。

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。

ある程度具体的・
状況を限定する方
が分析しやすい
(アンケートを
作成しやすい)

統計データ等の根拠も
探してみよう

アンケートは明確かつ限定的にして

データを問題の発見・解決に
活用する方法

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のどっち？

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

どのように分析する
かも考えてみよう

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）

数値・自由記述・選択肢をからめて

全用甲アンケート (回答) 1/10/16

最終編集: 5分前

データを分析する

A	B	C	D	E
回答 No	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)	あなたは「勉強」にどのようなイメージを持っていますか？	あなたは毎日、十分に睡眠を取れていると思いますか？
1	250	300	時間に追われて勉強をするのは好きではないけど、自分から進んで今やりたいと思う勉強をするのは楽しいと思う。	全くそう思わない
2	60	480	社会で生きていくための知識をつける作業	とてもそう思う
3	180	300	煩業	そう思う
4	1	5	面倒	とてもそう思う
5	60	250	学生がやるべきことの一つ	そう思わない
6	120	360	つらい	とてもそう思う
7	120	360	時間の搾取	そう思う
8	0	360	勉強自体はめっちゃくちゃ嫌いってわけではないけど、勉強を押し付けられるのが嫌い。	そう思う
9	240	470	努力次第で結果が決まるもの	とてもそう思う
10	240	340	単位を取るため 定期試験の点数を取るため	そう思わない
11	120	360	すればするほどいい	そう思わない
12	90	380	平等	そう思わない
13	60	330	将来に役立つが、めんどくさいもの。	そう思わない
14	170	360	学びは大切だという理屈は理解できるが、努力をすることの辛さがあるのでマイナスなイメージを持っている。	そう思わない
15	30	360	やったほうが良いんだろうなあ	そう思わない
16	200	320	一つの意味について色んな使い方があること	そう思わない
17	210	360	大変だが、わかるとおもしろいもの	そう思う
18	150	450	今の自分にはわからない、新しい知識をどんどん得ていくこと。	そう思う

数値・自由記述・選択肢をからめて

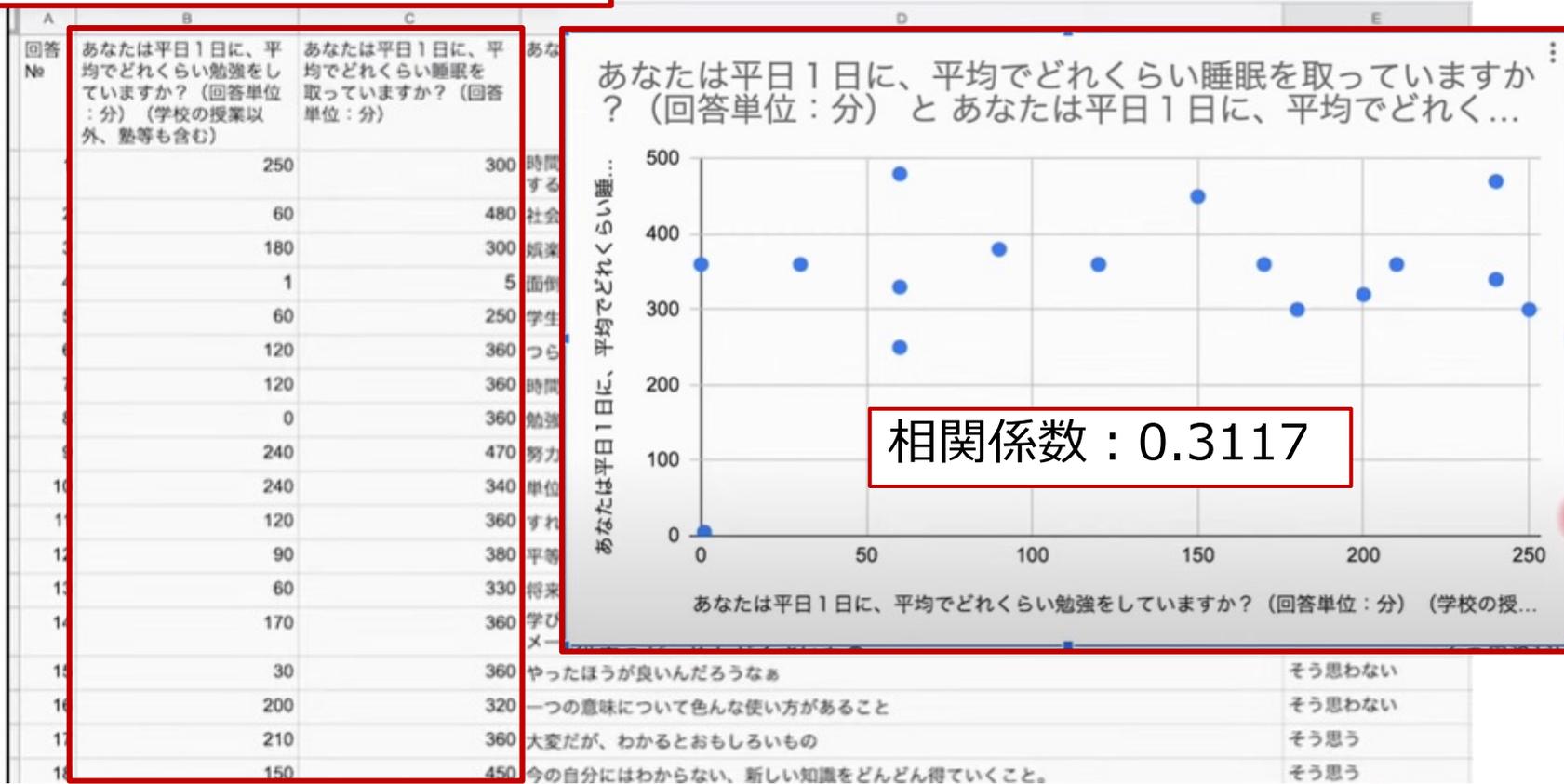
数値なら基本統計量(平均・分散・四分位点など)を出しやすい, 散布図も書きやすい

選択肢で聞くと層別にしやすい(割合がとりやすい)

A	B	C	D	E
回答 No	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)	あなたは「勉強」にどのようなイメージを持っていますか？	あなたは毎日、十分に睡眠を取れていると思いますか？
1	250	300	時間に追われて勉強をするのは好きではないけど、自分から進んで今やりたいと思う勉強をするのは楽しいと思う。	全くそう思わない
2	60	480	社会で生きていくための知識をつける作業	とてもそう思う
3	180	300	娯楽	そう思う
4	1	5	面倒	とてもそう思う
5	60	250	学生がやるべきことの一つ	そう思わない
6	120	360	つらい	とてもそう思う
7	120	360	時間の搾取	そう思う
8	0	360	勉強自体はめっちゃくちゃ嫌いってわけではないけど、勉強を押し付けられることが嫌い。	そう思う
9	240	470	努力次第で結果が決まるもの	とてもそう思う
10	240	340	単位を取るため 定期試験の点数を取るため	そう思わない
11	120	360	すればするほどいい	そう思わない
12	90	380	平等	そう思わない
13	60	330	将来に役立つが、めんどくさいもの。	そう思わない
14	170	360	学びは大切だという理屈は理解できるが、努力をすることの辛さがあるのでマイナスなイメージを持っている。	そう思わない
15	30	360	やったほうが良いんだろうなあ	そう思わない
16	200	320	一つの意味について色んな使い方があること	そう思わない
17	210	360	大変だが、わかるとおもしろいもの	そう思う
18	150	450	今の自分にはわからない、新しい知識をどんどん得ていくこと。	そう思う

数値・自由記述・選択肢をからめて

数値なら基本統計量(平均・分散・四分位点など)を出しやすい, 散布図も書きやすい

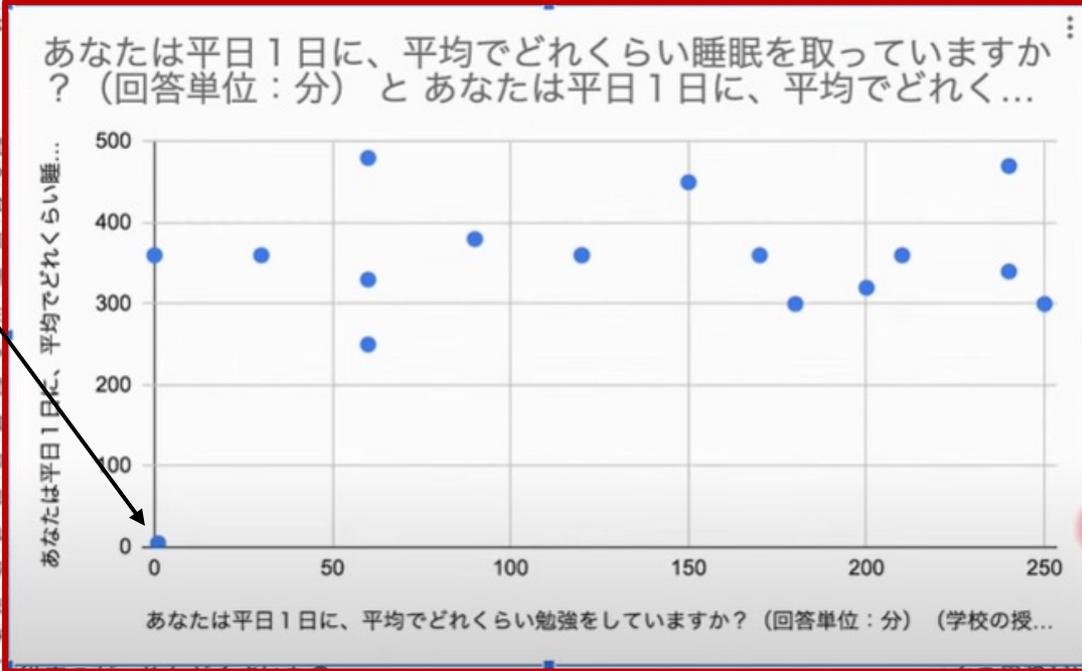


数値・自由記述・選択肢をからめて

数値なら基本統計量(平均・分散・四分位点など)を出しやすい, 散布図も書きやすい

回答 No	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)	あなた
1	250	300	時間
2	60	480	社会
3	180	300	卒業
4	1	5	面倒
5	120	250	学生
6	360	360	つら
7	360	360	時間
8	0	360	勉強
9	240	470	努力
10	240	340	単位
11	120	360	すれ
12	90	380	平等
13	60	330	将来
14	170	360	学び
15	30	360	やったほうが良いんだろうなあ
16	200	320	一つの意味について色んな使い方があること
17	210	360	大変だが、わかるとおもしろいもの
18	150	450	今の自分にはわからない、新しい知識をどんどん得ていくこと。

外れ値の発見



360	360	やったほうが良いんだろうなあ	そう思わない
320	320	一つの意味について色んな使い方があること	そう思わない
360	360	大変だが、わかるとおもしろいもの	そう思う
450	450	今の自分にはわからない、新しい知識をどんどん得ていくこと。	そう思う

数値・自由記述・選択肢をからめて

数値なら基本統計量(平均・分散・四分位点など)を出しやすい, 散布図も書きやすい

A	B	C	D	E
回答 No	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）	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）とあなたは平日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）
1	250	300	時間勉強するの	
2	60	480	社会	
3	180	300	娯楽	
4	1	5	面倒	
5	60	250	学生が	
6	120	360	つらい	
7	120	360	時間勉強	
8	0	360	勉強	
9	240	470	努力	
10	240	340	単位	
11	120	360	すれ	
12	90	380	平等	
13	60	330	将来	
14	170	360	学びは大切なという理由は理解できるが、努力をすることの辛さがあるのでマイナスなイメージを持っている。	そう思わない
15	30	360	やったほうが良いんだらうなぁ	そう思わない
16	200	320	一つの意味について色々な使い方があること	そう思わない
17	210	360	大変だが、わかるとおもしろいもの	そう思う
18	150	450	今の自分にはわからない、新しい知識をどんどん得ていくこと。	そう思う

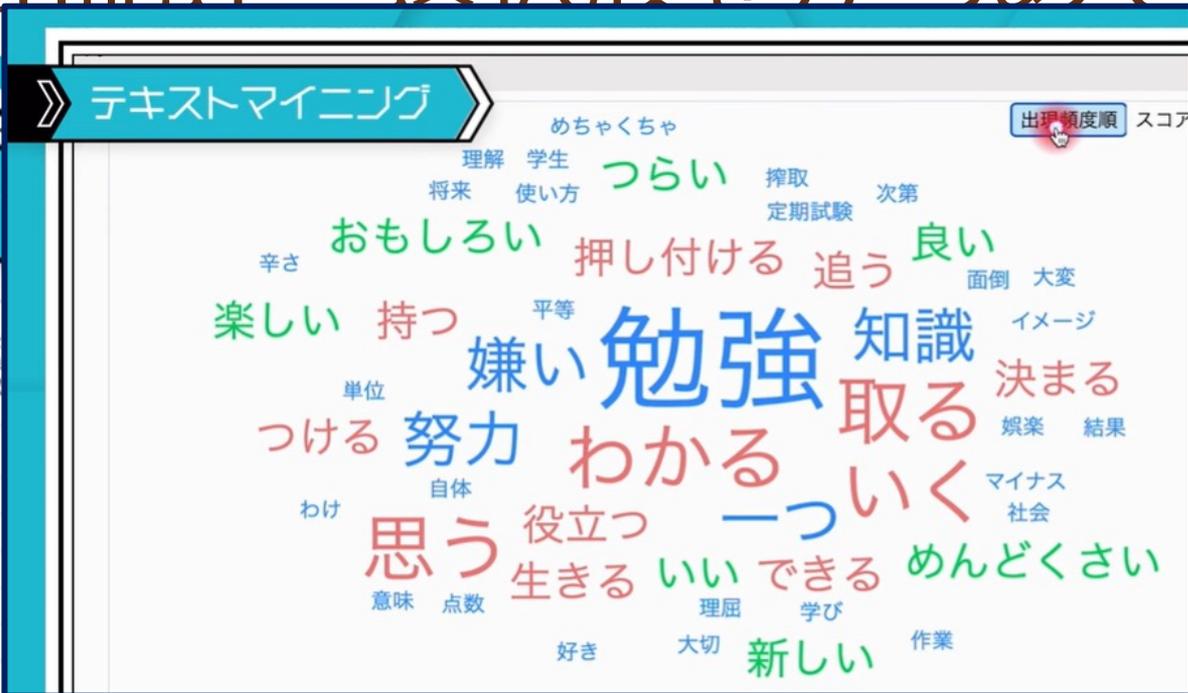


数値・自由記述・選択肢をからめて

データ

全用アサケート (回答)

テキストマイニング



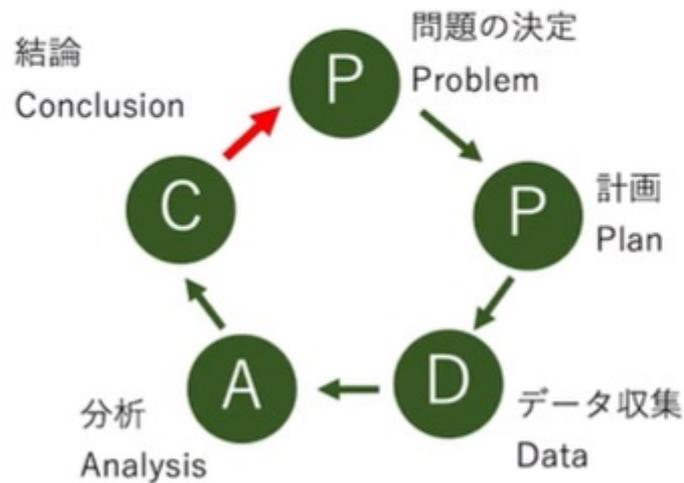
出題頻度順 スコア

A	B		
回答 No	あなたは平日1日に、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)		
1			
2			
3			
4			
5			
6			
7			
8	0	360 勉強自体はめっちゃ嫌いってわけではないけど、勉強を押し付けられるのが嫌い。	そう思う
9	240	470 努力次第で結果が決まるもの	とてもそう思う
10	240	340 単位を取るため 定期試験の点数を取るため	そう思わない
11	120	360 すればするほどいい	そう思わない
12	90	380 平等	そう思わない
13	60	330 将来に役立つが、めんどくさいもの。	そう思わない
14	170	360 学びは大切だという理屈は理解できるが、努力をすることの辛さがあるのでマイナスなイメージを持っている。	そう思わない
15	30	360 やったほうが良いんだろうなあ	そう思わない
16	200	320 一つの意味について色んな使い方があること	そう思わない
17	210	360 大変だが、わかるとおもしろいもの	そう思う
18	150	450 今の自分にはわからない、新しい知識をどんどん得ていくこと。	そう思う

分析から次の分析へ

データを問題の発見・解決に
活用する方法

PPDACサイクル

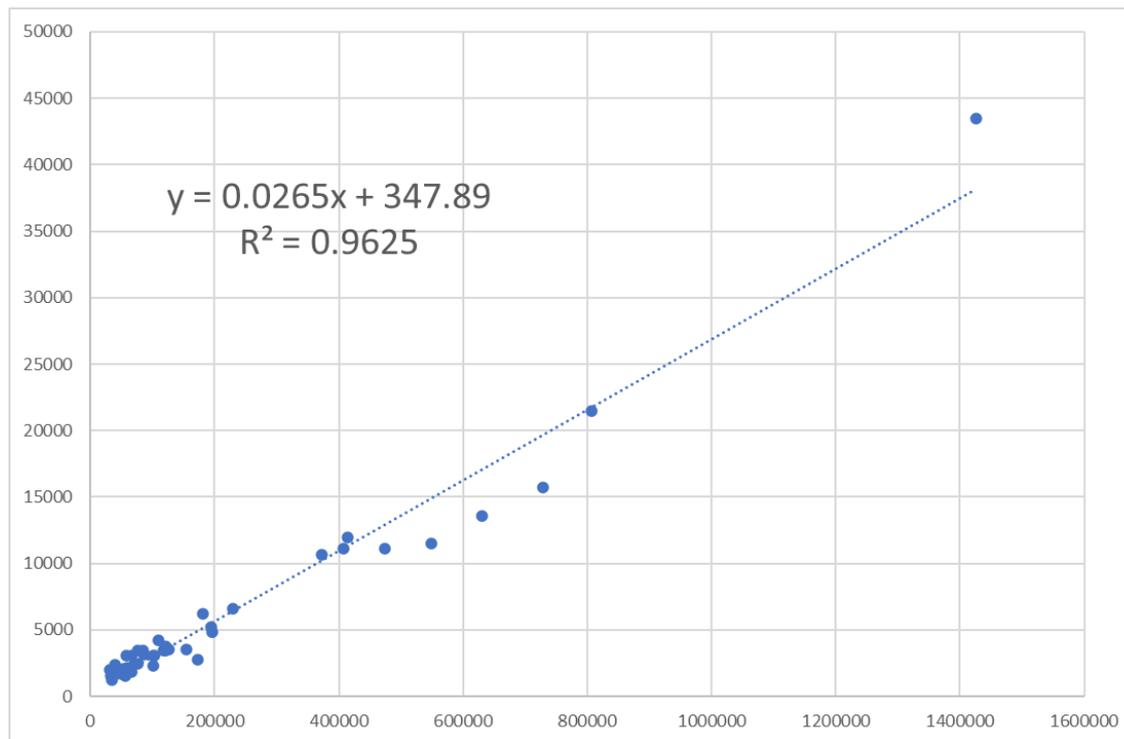


(例)

- 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。

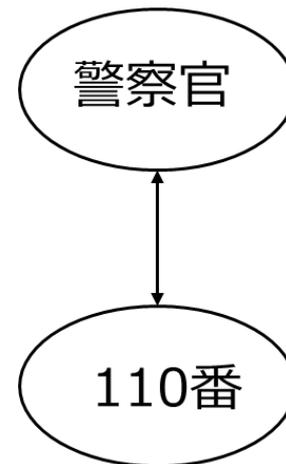
警察官と通報件数は相関関係？

- 各都道府県の警察官の人数(縦軸)と110番通報件数(横軸)の
関係を見ると…



警察官と通報件数は相関関係？

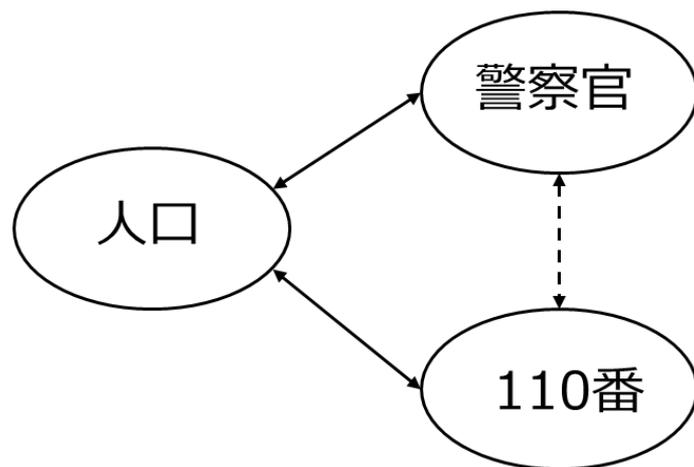
- 各都道府県の警察官の人数(縦軸)と110番通報件数(横軸)の
関係を見ると…
→ 警察官が多いと110番通報件数が多い(治安が悪くなる)
- 本当に??



警察官と通報件数は相関関係？

- 各都道府県の警察官の人数(縦軸)と110番通報件数(横軸)の
関係を見ると…
→ 警察官が多いと110番通報件数が多い(治安が悪くなる)

- 本当に??



- 擬似相関の有名な例はいくつかあるので、それを利用して
授業展開するのもよい

いただいた質問内容から…②

- **授業時数が足りませんどうしたらよいですか？**

→ オープンデータを用いれば，データ収集部分の時間は省略可能

(他にも工夫してみると…)

- 知識を確認するだけの小テストのようなものは，与えたプログラムを実行して確認するだけでよいものなどは，生徒の都合が良い時間で取り組むようにする
- PPDACサイクルを回しながら理解を深める流れをつくる (生徒がPPDACサイクルをまわす時間が多くなるように)
- データを深掘りするきっかけまでは授業で，実際に深掘りする試行錯誤は自宅で行うようにする

→ それでも時間が足りない場合は…

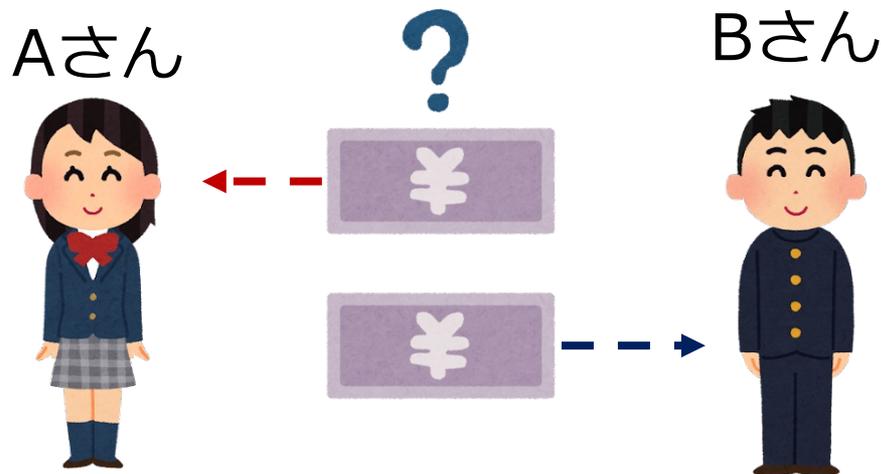
実験をふまえたデータ解析もあり

- 2人組を作る(Aさん, Bさん)
- まず, Aさんに1万円渡す(渡す状況を想定する)



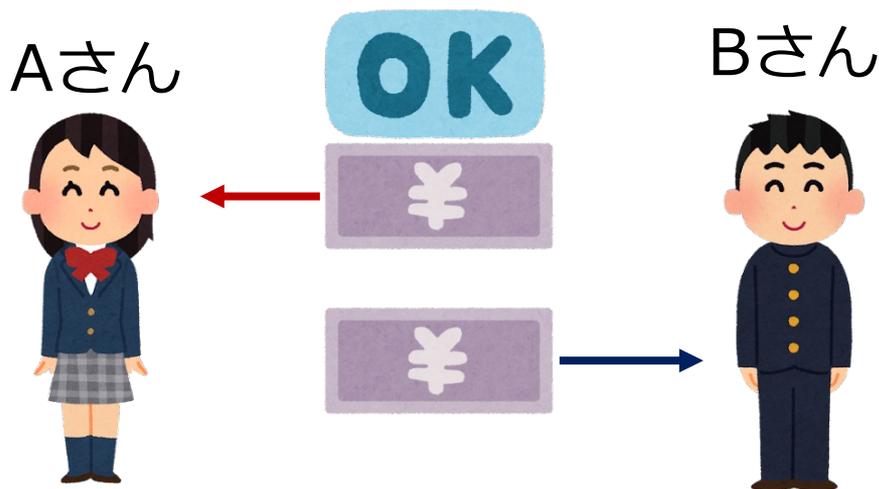
実験をふまえたデータ解析もあり

- 2人組を作る(Aさん, Bさん)
- まず, Aさんに1万円渡す(渡す状況を想定する)
- Aさんは, 自分とBさんとの1万円の配分を提案する
(例えばAさん5000円, Bさん5000円)



実験をふまえたデータ解析もあり

- 2人組を作る(Aさん, Bさん)
- まず, Aさんに1万円渡す(渡す状況を想定する)
- Aさんは, 自分とBさんとの1万円の配分を提案する
(例えばAさん5000円, Bさん5000円)
 - BさんがAさんの提案にOKすれば, Aさんの配分提案どおりのお金が, AさんBさんそれぞれもらえる



実験をふまえたデータ解析もあり

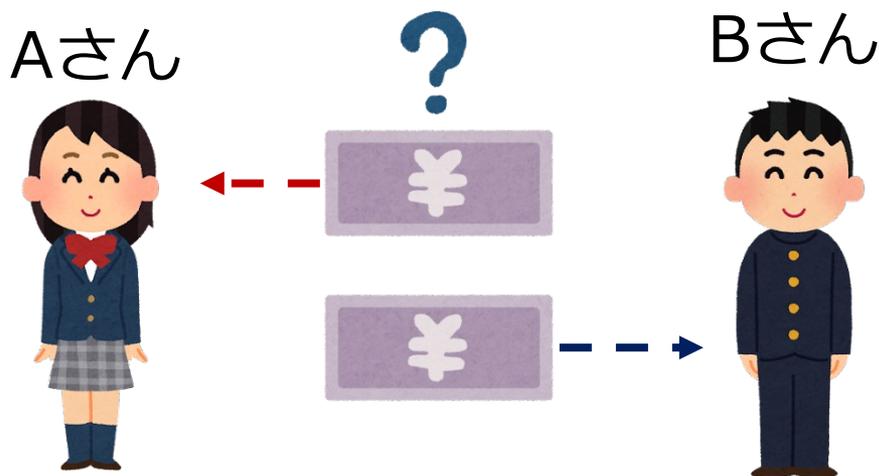
- 2人組を作る(Aさん, Bさん)
- まず, Aさんに1万円渡す(渡す状況を想定する)
- Aさんは, 自分とBさんとの1万円の配分を提案する(例えばAさん5000円, Bさん5000円)
 - Bさんが拒否すれば, 両方とも, 何ももらえない(1万円は没収されます)



実験をふまえたデータ解析もあり

- 2人組を作る(Aさん, Bさん)
- まず, Aさんに1万円渡す(渡す状況を想定する)
- Aさんは, 自分とBさんとの1万円の配分を提案する
(例えばAさん5000円, Bさん5000円)

Q : Aさんの最も良い提案は何でしょうか？



授業では…

- Aさんが提案した配分額とその結果を記録
- NGだった場合, いくらだったらOKとしたかも記録
- これを他のメンバーで組み替えて実施する

- 結果をまとめる(そのクラスの配分提示額の平均や分散, 受諾率, 受け入れ可能額の平均や分散など)
- 実際に合理的とされる結果を示す
(おそらくその結果は納得されない可能性が大きい)

- 他の国々でやって結果なども示し, 自分たちのクラスがどうだったか, なぜそうなったのかの分析を行う

ちなみに結果は…

解答：理論上最も良い提案(最良提案)は…,

『Aさん：9999円， Bさん：1円』

(理由)

- Bさんはどんな提案であろうとも，提案にOKすれば，拒否する(=0円)よりはプラスになるので，Bさんにプラスを与えつつ，Aさんが最大限プラスになる提案をするはず！

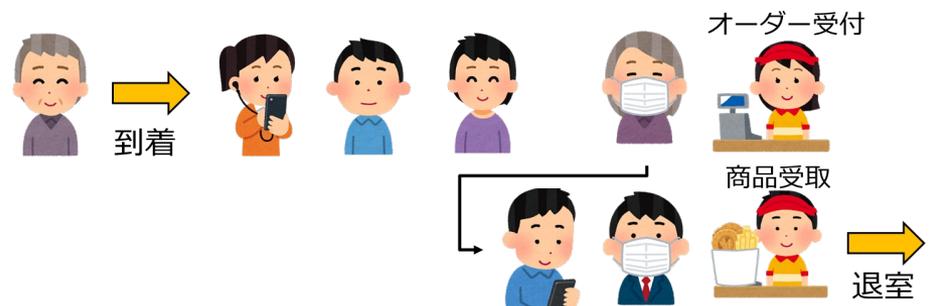
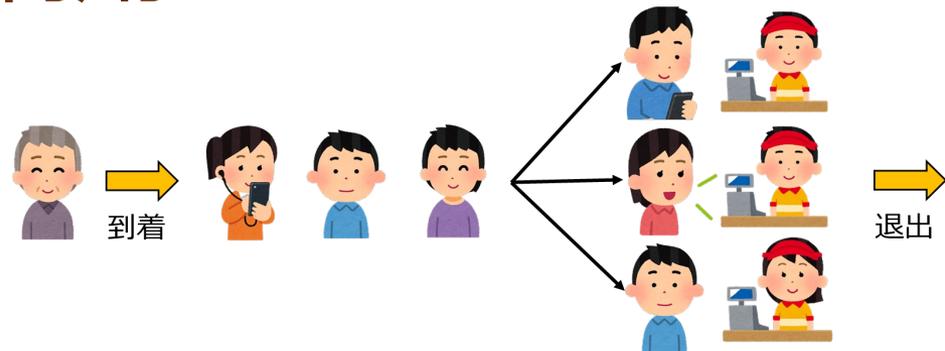
けれども…

- 結果はおそらく，4000円～5000円くらいになる(はず)
- なぜ理論とは異なる結果になるのだろうか？
- 今回は1万円だったが，1000円で実施した場合，10万円で実施した場合などでは変化するのだろうか？

(参考) 小林佳世子, "最後通牒ゲームの謎 ◇進化心理学からみた行動ゲーム理論入門", 日本評論社 (2021)

シミュレーションの利用

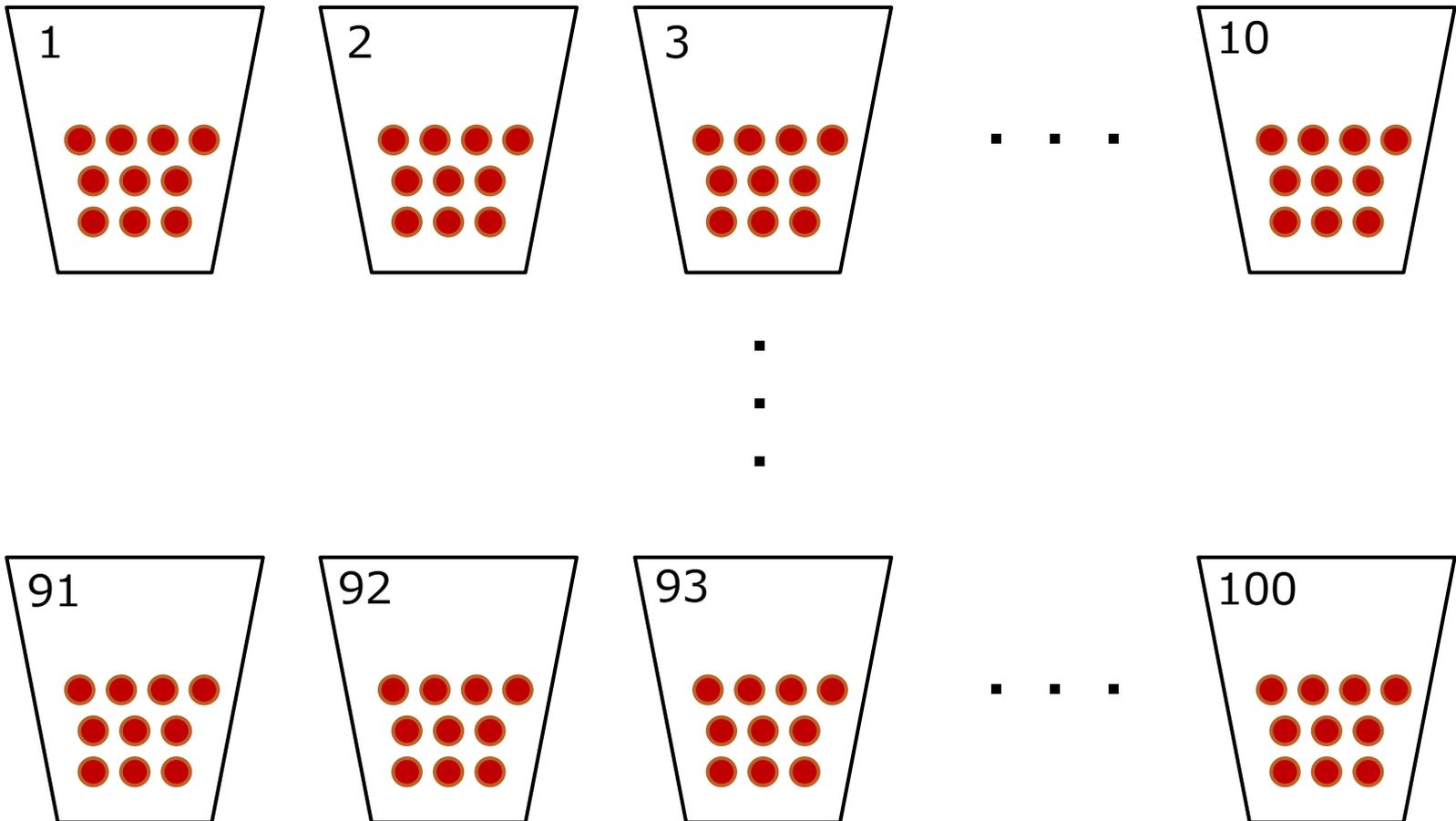
- どれが最も行列の長さを短くできる?
or 待ち時間が短い?
- (事前にシミュレーション環境を整えたうえで)到着間隔やサービス時間の様々なパターンで分析(学校の食堂などで調査)
- カウンター数も考慮
- (後述するが)シミュレーションとデータ分析をうまくつなげる



(参考)

データサイエンスにはやはり数学が必要？

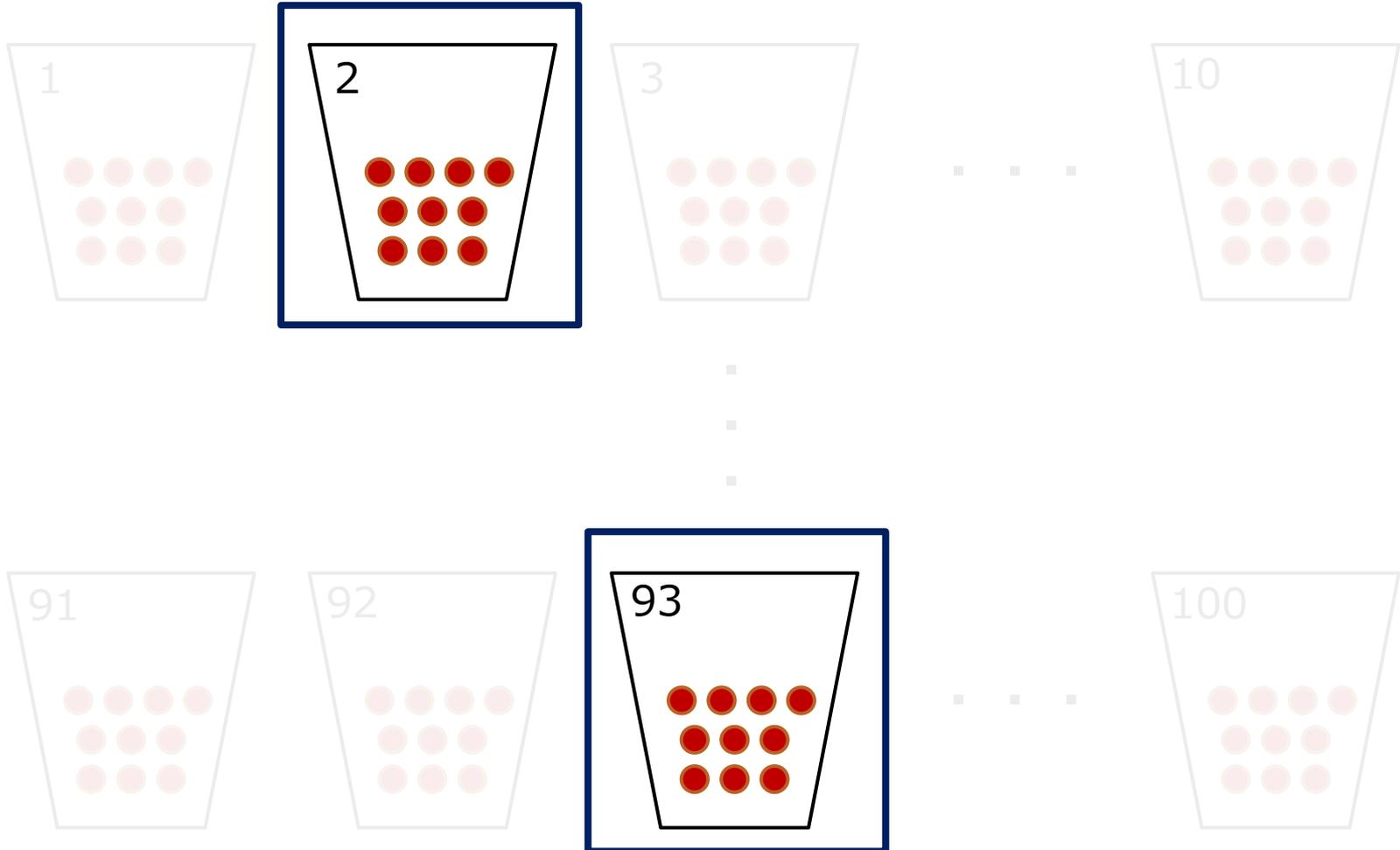
- 1カップあたり10個の赤球が入っているカップが100個



(参考)

データサイエンスにはやはり数学が必要？

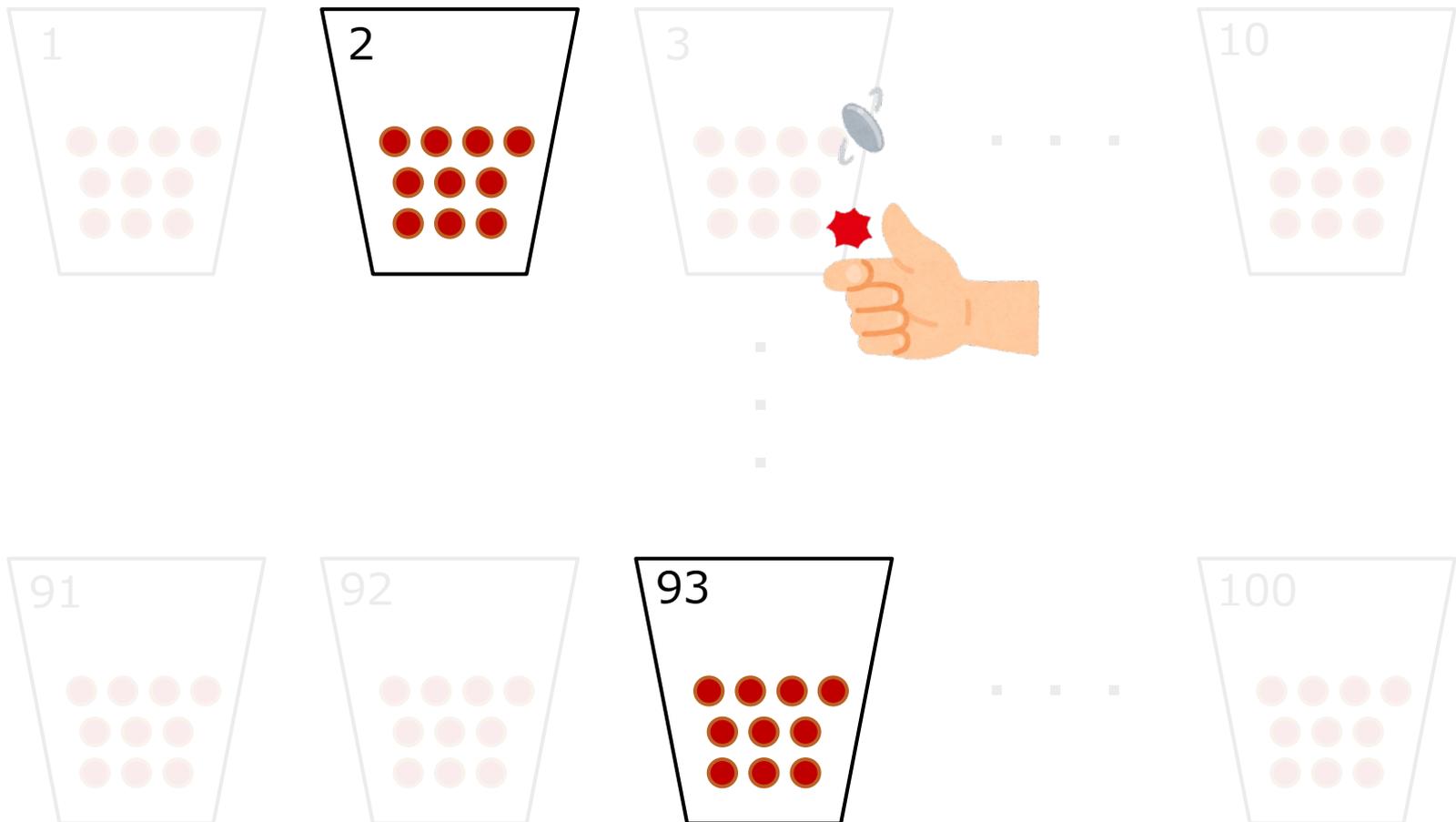
- 2つのカップをランダムに選択



(参考)

データサイエンスにはやはり数学が必要？

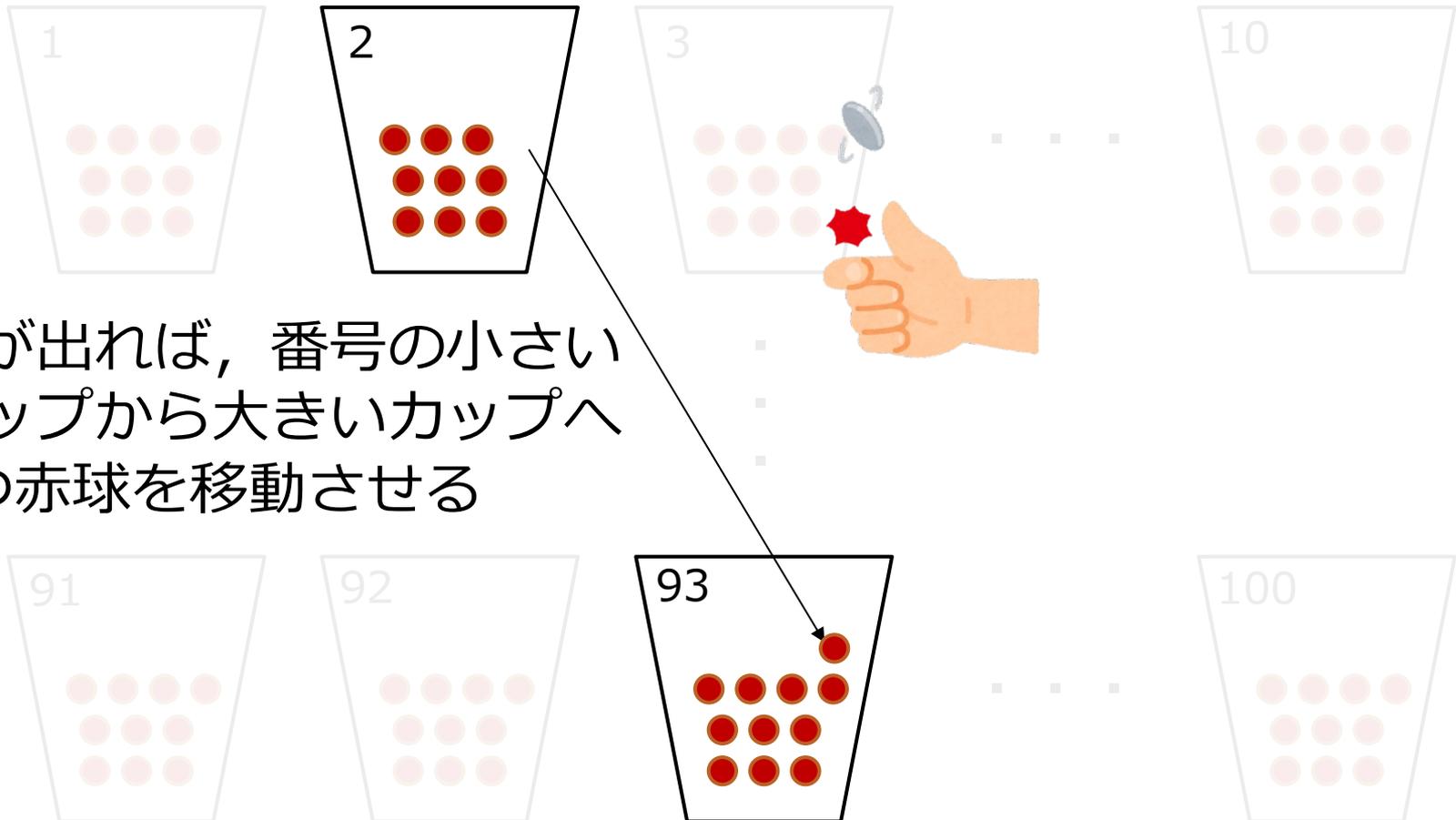
- 表裏が同じ確率で出るコインを1枚投げる



(参考)

データサイエンスにはやはり数学が必要？

- 表裏が同じ確率で出るコインを1枚投げる

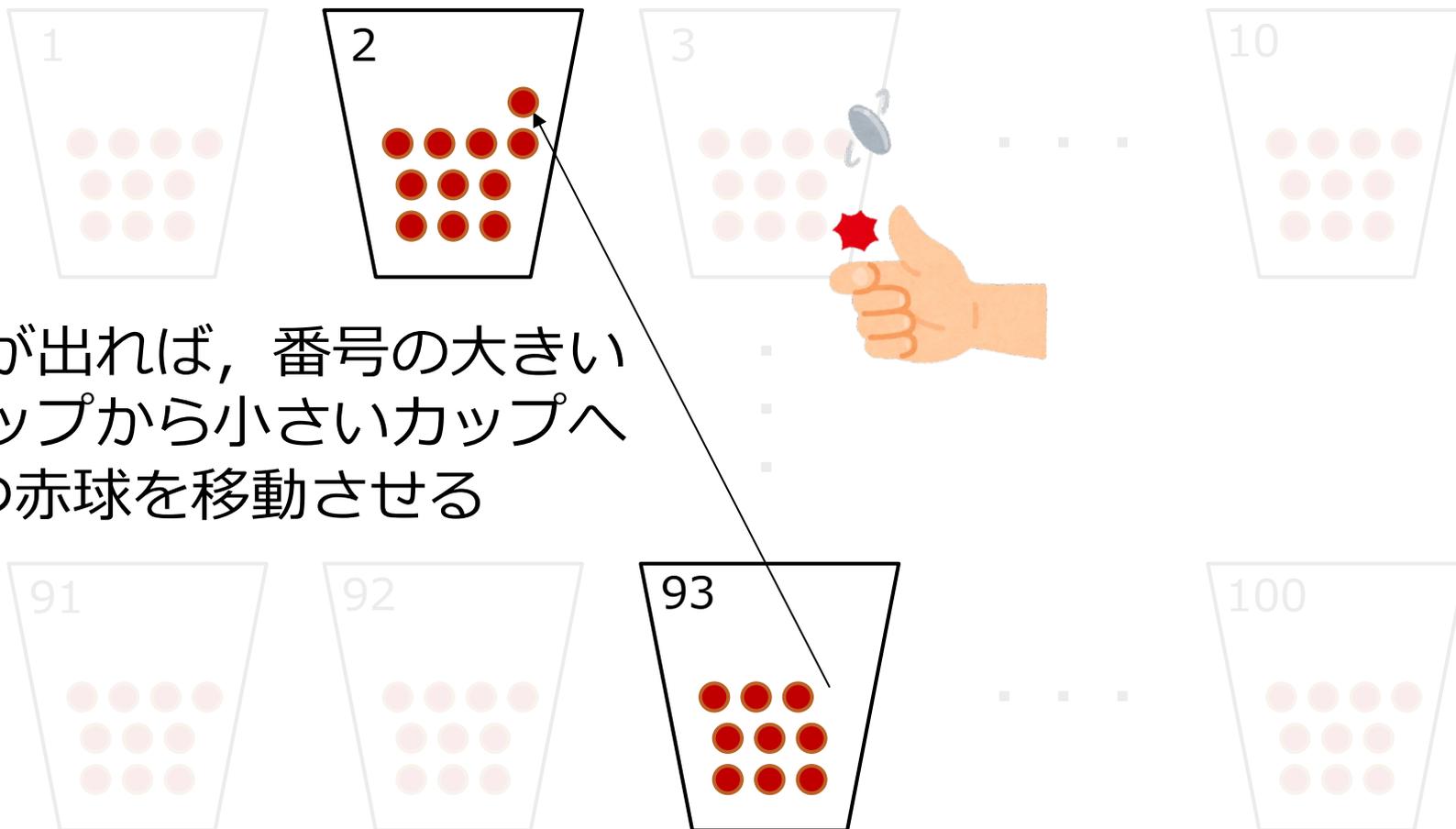


表が出れば，番号の小さい
カップから大きいカップへ
1つ赤球を移動させる

(参考)

データサイエンスにはやはり数学が必要？

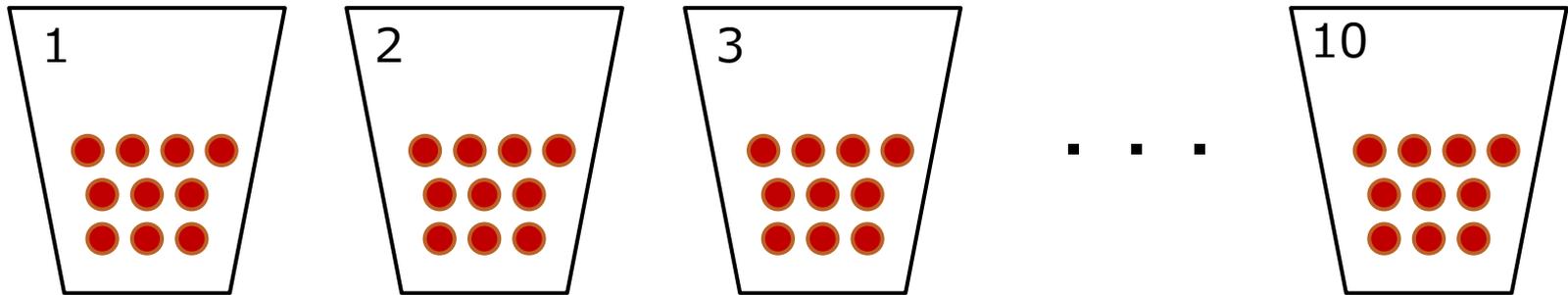
- 表裏が同じ確率で出るコインを1枚投げる



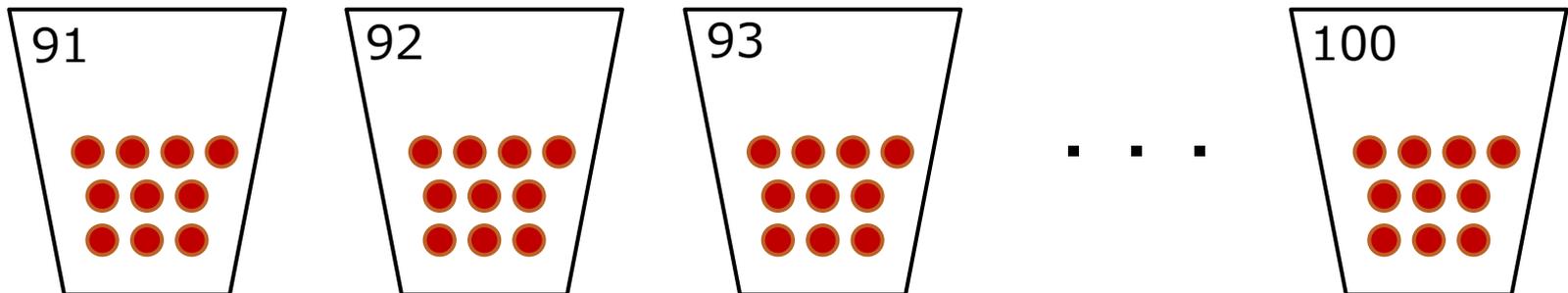
裏が出れば、番号の大きい
カップから小さいカップへ
1つ赤球を移動させる

考えてみましょう！

- これをめちゃくちゃ多くの回数(数千万回)繰り返す

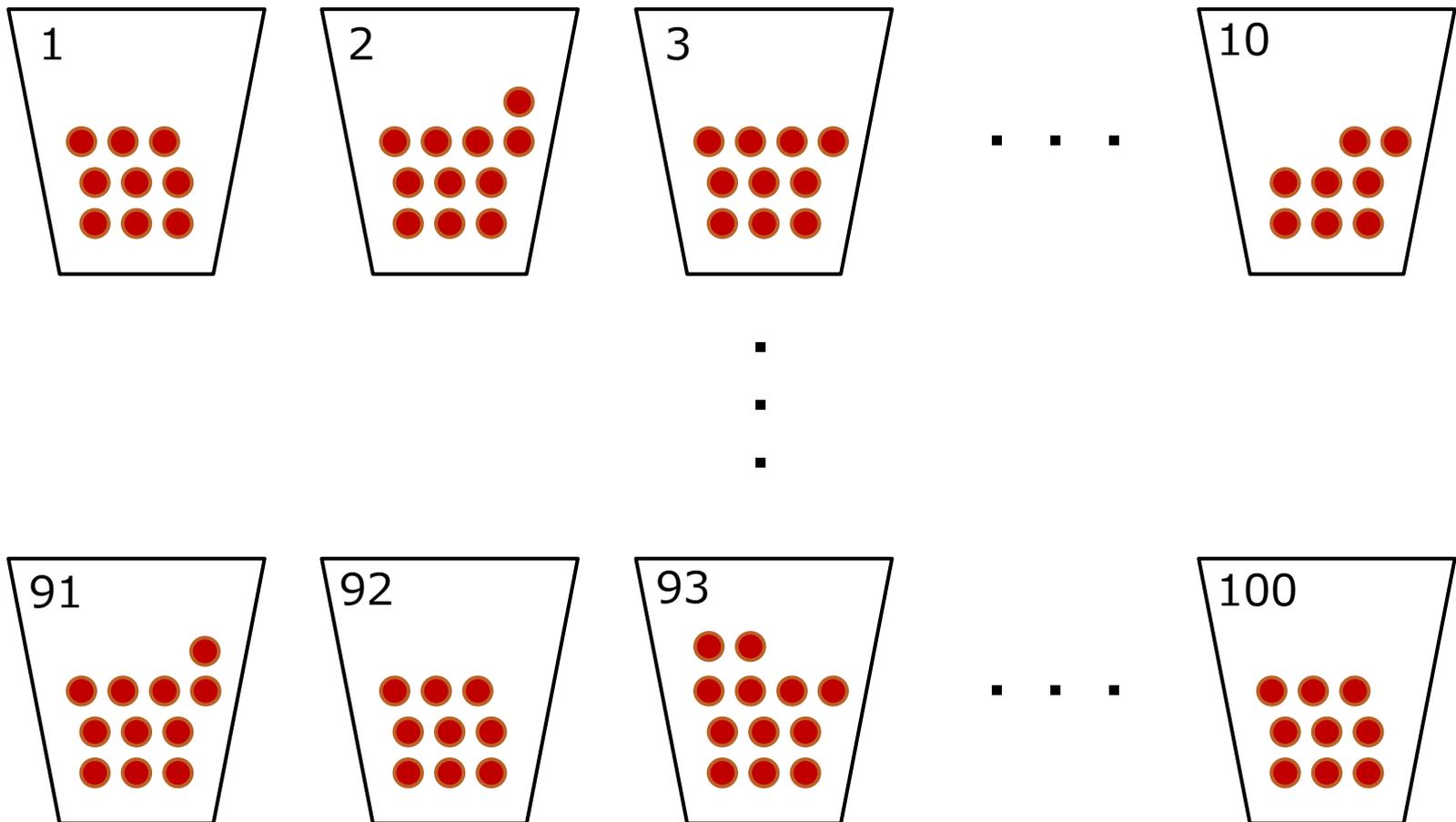


Q：繰り返した後，カップの状況はどうなるでしょうか？
次の2つの選択肢から選んでください



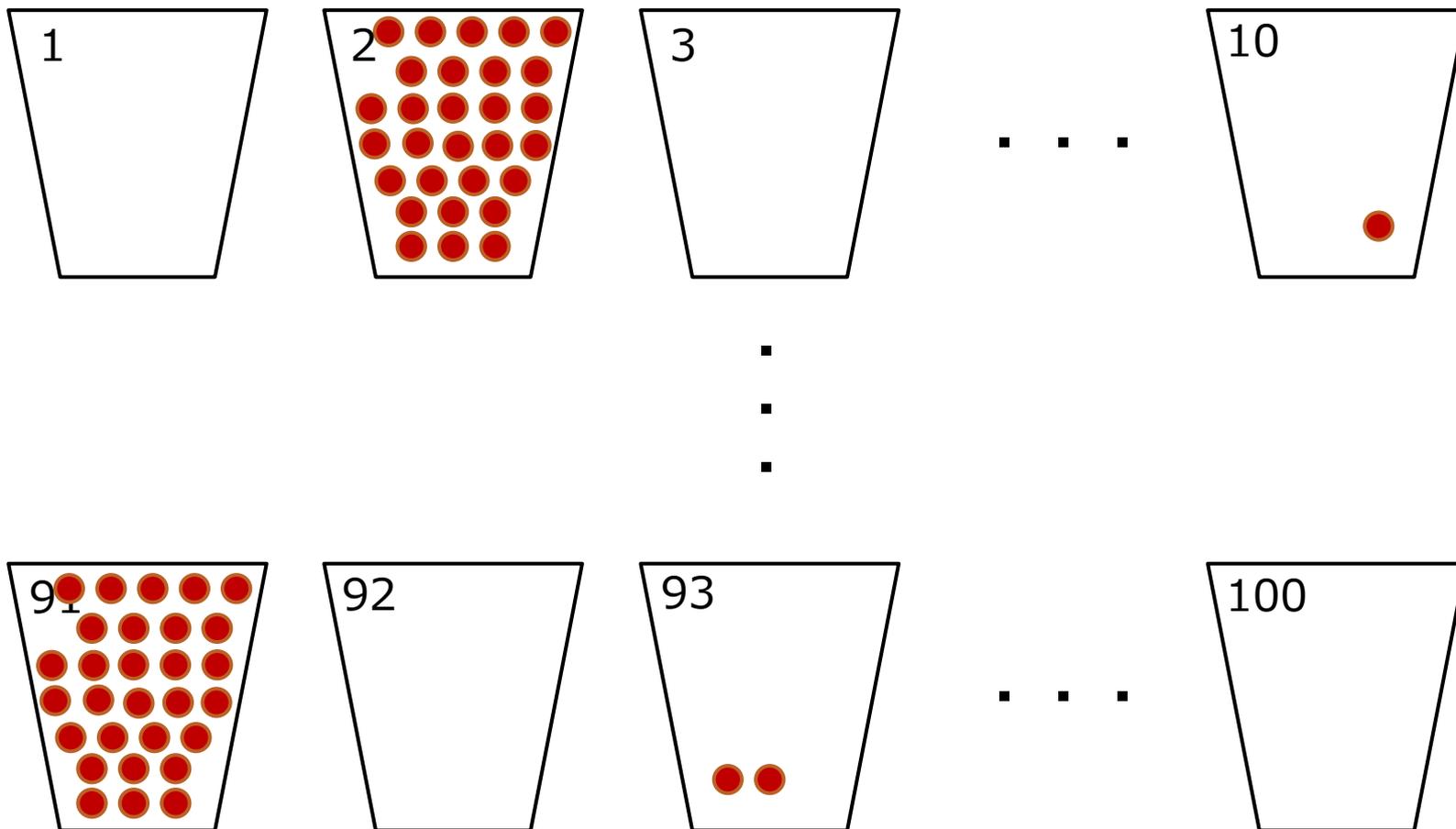
考えてみましょう！

- A : すべてのカップにおおよそ同じくらいの赤球がある



考えてみましょう！

- B : いくつかのカップに赤球が集中し, 空のカップも多数



この問題をまとめると

- 1カップあたり10個の赤球が入っているカップが100個
- 2つのカップをランダムに選択
- 表裏が同じ確率で出るコインを1枚投げる
 - 表が出れば、番号の小さいカップから大きいカップへ1つ赤球を移動させる
 - 裏が出れば、番号の大きいカップから小さいカップへ1つ赤球を移動させる
- これをめちゃくちゃ多くの回数繰り返した後、カップの状況はどうなるでしょうか？
 - A：すべてのカップにおおよそ同じくらいの赤球がある
 - B：いくつかのカップに赤球が集中し、空のカップも多数

ちなみに結果は…

- これをめちゃくちゃ多くの回数繰り返した後、カップの状況はどうなるでしょうか？
 - A：すべてのカップにおおよそ同じくらいの赤球がある
 - **B**：いくつかのカップに赤球が集中し、空のカップも多数
- シミュレーションは、「コンピュータとプログラミング」で扱うため、シミュレーション結果を「データの活用」にどのようにつなげていくかは要検討

いただいた質問内容から…③

- **実社会や大学進学後を見据えた内容を展開するのに適切なデータ等は？**
 - 前述のオープンデータでもOK. スポーツデータも分析も多岐で行うことができ、かつ身近なのでとっつきやすい
- マーケティング系のデータを扱うことができればよいが、先端の機械学習・AI手法が利用する可能性が高い…
(もしくはデータをオープンにしてくれない場合も多い…)
- 「国勢調査」と「経済センサス」のデータを利用すれば、ある程度の分析は可能

Mapから読み解くマーケティング

- 「コンビニエンスストアやコーヒーショップって隣り合う場所に合ったりするけど、どうしてなんだろう？」
- 実際にコンビニエンスストアやコーヒーショップが隣接している(もしくはかなり近い場所にある)場所がどのくらいあるか集計する
- 自社の他の店舗, 同業種の競合他社, ドラッグストアなどの競合他社との距離関係も見てみると面白いかも
- 参考: ホテリングの立地モデル(1929年発表)
<http://gametheory.jp/page.php?id=96>
(最終閲覧日: 2022年2月22日)



例：スポーツにおける都市伝説？

- 素朴な疑問：スポーツ競技でよく「今、流れが来てます」という言葉を耳にするけど、本当に「流れ」ってある？
- 野球であれば、それなりにデータが存在する
(例えば、日刊スポーツのWebページより
<https://www.nikkansports.com/baseball/highschool/archives.html>)
- 「四球を出すと流れが悪くなって、失点しやすくなる？」
- 「スクイズ作戦に意味はあるの？」
- 「ジョックロックは本当に魔曲？」

例：スポーツにおける都市伝説？

- 例えば，高校野球(甲子園の春・夏の大会)における6回以降の安打出塁後と四球後の投球コースの違い

コース	安打後		四死球後	
	投球数(球)	投球割合	投球数(球)	投球割合
高	88	0.188	55	0.176
中	157	0.335	134	0.428
低	223	0.476	124	0.396

- 真ん中や低めに来やすい
 - 叩きつけるバッティングでゴロを転がしやすい
 - ヒットエンドランをかけやすい？
- スクイズ作戦の有効性に関する研究なども，いろいろと発表されている

サンプル問題がスポーツのデータサイエンス

第3問 次の文章を読み、後の問い(問1～4)に答えよ。

仮説を立てる

S高等学校サッカー部のマネージャーをしている鈴木さんは、「強いサッカーチームと弱いサッカーチームの違いはどこにあるのか」というテーマについて研究している。鈴木さんは、ある年のサッカーのワールドカップにおいて、予選で敗退したチーム(予選敗退チーム)と、予選を通過し、決勝トーナメントに進出したチーム(決勝進出チーム)との違いを、データに基づいて分析することにした。このデータで各国の代表の32チームの中で、決勝進出チームは16チーム、予選敗退チームは16チームであった。

比較のため、全体を2つに分割(層別)

例：文章の解析もいい題材

- 動画共有プラットフォームで「バズりやすいタイトルってあるの？」 「再生回数が伸びやすい動画ってどんな動画？」
- 売れている小説・音楽の歌詞は時代を反映してる？
- 作家・作詞家ごとに特徴(例：恋愛系の言葉が好き, ある特定の季節の作品が多い など)がある？

- 文章の中に出てくる「名詞・形容詞・副詞」の頻度を比較するだけでも, ある程度分析できる
- データの例：(文学系なら)青空文庫
<https://www.aozora.gr.jp/>
- (API等を使いこなせる先生・学生がいれば, 動画共有プラットフォームなどからデータを収集, まとめるのもあり)

例：文章の解析もいい題材

- 動画共有プラットフォームで「バズりやすいタイトルってあるの？」 「再生回数が伸びやすい動画ってどんな動画？」
- 売れている小説・音楽の歌詞は時代を反映してる？
- 作家・作詞家ごとに特徴(例：恋愛系の言葉が好き，ある特定の季節の作品が多い など)がある？

- 文章の中に出てくる「名詞・形容詞・副詞」の頻度を比較するだけでも，ある程度分析できる
- データの例：(文学系なら)青空文庫
<https://www.aozora.gr.jp/>
- (API等を使いこなせる先生・学生がいれば，動画共有プラットフォームなどからデータを収集，まとめるのもあり)

まとめ

- おそらく、前回の学習指導要領からの大きな変更点がこの「**データ分析**」「**データを用いた問題解決**」
 - まずは数値データ、構造化データから進めていくことが良いかと思えます(データ・手法ともに扱いやすい)
 - オープンデータを利用すれば、データ分析が始めやすいです
 - アンケートをとることで、PPDACを回すことができ、共通テストへのつながりも作りやすいと思えます
- データ分析は怖いものではありません！
難しい部分もあるかもしれませんが、
失敗を恐れずにトライしてみましよう！



(参考)分析のための概略表

- 分析対象のデータ(質的・量的)の比較対象と可視化方法

比較対象 \ 分析対象	なし(分析対象のみ)	質的データ	量的データ
質的データ	棒グラフ 円グラフ	分割表 (クロス集計) 層別	
量的データ	ヒストグラム 棒グラフ 円グラフ 折れ線グラフ	層別	散布図 相関係数