

---

# 情報 I No. 22

## データの活用①

---

年	2	組		番		名前	
---	---	---	--	---	--	----	--

第4章情報通信ネットワークとデータの活用 3節 データの活用

1. データの収集と整理 2. 数値データの分析 (教 P192-P195)

☞問題解決にデータ活用が有効であることを理解しよう。

【TRY】日本一〇〇を食べる都道府県を調べよう。その理由も考えよう。

〇〇に入れる食べ物	
ランキング 1～3位	
なぜ多いのか その理由は？	

【実習】 様々な都道府県別統計が紹介されている「とどらん」にあるデータを使って、データ同士の関係の深さ（相関関係）を調べよう。

【実習1】 QRコードを読み込み「とどらん」(<https://todo-ran.com/>)へ移動する。

※「とどらん」=さまざまなデータ（国土・インフラ、社会・政治、文化・暮らし・健康、娯楽・スポーツ、店舗分布）について、都道府県別のデータを公開



◇手順1 色々なデータの中から「なぜそうなのか」と自分が興味を持ったデータを3つ書き出そう

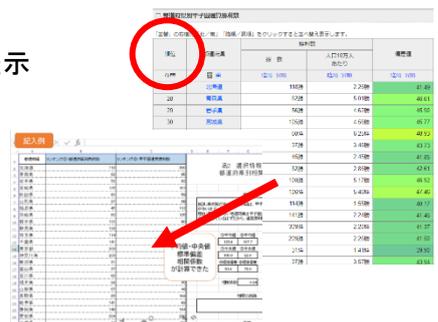
◇手順2 手順1で選んだデータ（それ以外でもよい）について、仮説（こうだからこれが増える）を立て、検証ができそうなデータを「とどらん」から探そう。候補は2つ考えよう。

仮説	(例)「高校数が多い都道府県」ほど「甲子園の勝利数」が多い	
	①	②
使うデータ	(例)「都道府県別高校数」と「都道府県別通算甲子園勝利数」	
	①	②
理由	(例) 予選で戦う回数が多いはずだから強いチームが甲子園に残るから	
	①	②

【実習2】 「とどらん」のデータをワークシートに入力し、相関係数を算出する。(Excelを使用)

◇手順1

- ①使うデータの一覧表（地図をスクロールさせた下にある）を表示
- ②表の「北」ボタンを押し、データを北からの順番に並べる
- ③表の「北海道の順位～沖縄の偏差値まで選択してコピーする
- ④自分のExcelのワークシートのシートにコピーする。
- ⑤単位が入っているデータは「置換」を使って数字のみにする



◇手順 2

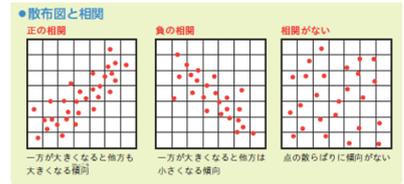
①自動的に計算された平均値、中央値、標準偏差、相関係数を記録する。

	データ①	データ②
①平均値（データ全体の平均）		
②中央値（真ん中の県のデータ）		
③標準偏差（データの散らばり・幅）		
④相関係数（2つのデータの関係ありなし）		

【実習 3】仮説（2つのデータの相関関係）が正しかったか、検証しよう。

◇手順 1 相関係数をもとに、次の表で相関関係があるかどうか判定しよう。

相関係数	判定
-1~-0.7	強い負の相関がある
-0.7~-0.5	負の相関がある
-0.5~-0.3	弱い負の相関がある
-0.3~0.3	相関はない
0.3~0.5	弱い正の相関がある
0.5~0.7	正の相関がある
0.7~1	強い正の相関がある



相関関係の判定	(例) 強い負の相関がある
因果関係の判定	
考察	

※考察で書くこと ・自分の仮説が正しかったか、違ったのか。  
 ・なぜそういう結果となるのか、自分なりの分析を書く

【知識の整理】

1 データの活用

・ ( ) = データをコンピュータや統計学など専門的な知識を使って分析する学問  
 → データを分析すれば未来を予測できる (例) コンビニの売上データ → 未来の売上の予測



・ データサイエンスの分析対象  
 従来の分析対象 ( 構造化データ ) = 数値や記号など表形式に整理できるデータのこと  
 → 現在は ( 非構造化データ ) = 音声や動画、画像も分析の対象になっている



・ ( ) = 大量かつ多様で発生頻度の高いデータ (例) SNS の書込み、センサなど  
 → 分析すればさまざまなことを予測できる (例) SNS の書き込みデータ → 何が売れるか予測

## 2 データの分析

### ①データの収集

- ・ ( ) = 多くの人に同じ質問を行い、多くの回答を集める方法
  - ・ 方法による分類：(質問紙調査)、(Web 調査)、(街頭調査) = 対面で回答を得る
  - ・ 対象による分類：(全数調査) = 全員が対象、(標本調査) = 一部を抽出して調べる方法
- ・ ( ) の利用 = 国や自治体、研究・教育機関、企業などが広く公開するデータ



### ②データの整理

- 1) データの確認：欠けてる値 (欠損値)、他と比べて大きく外れた値 (外れ値) を除く
- 2) 集計する：項目ごとに計算する (単純集計)、2 以上のカテゴリ間で集計する (クロス集計)

## 3 データの種類 …データの種類によって計算できる・できないがある

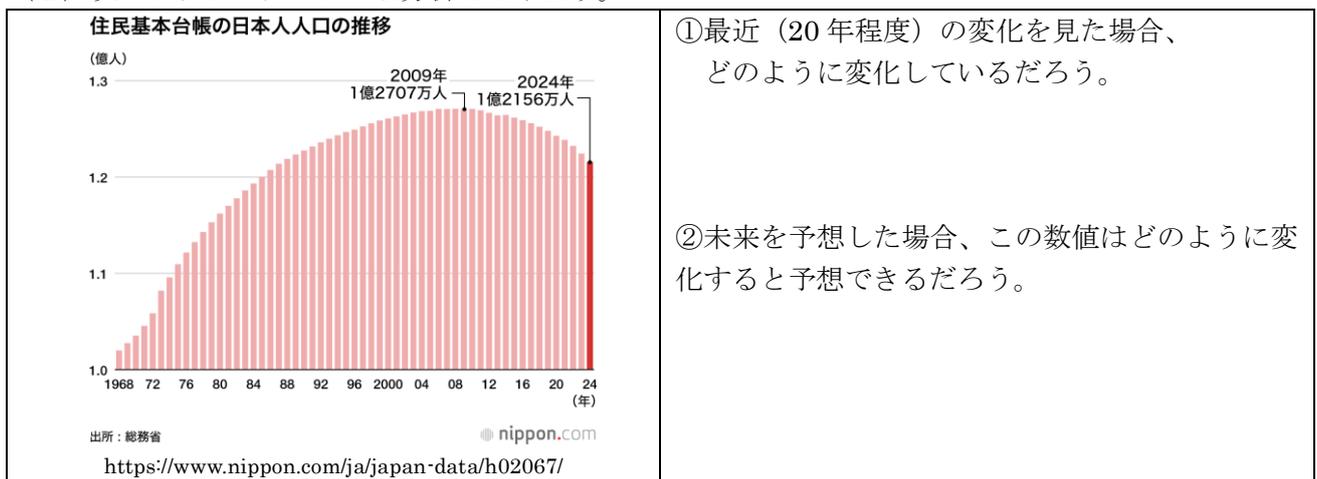
( ) = 数値でも計算できない	(名義尺度)	分類のため数値を割り当てた (例) 性別、血液型
	(順序尺度)	数値の順序に意味がある (例) 5 段階評価、おみくじ
( ) = 数値で計算ができる	(間隔尺度)	0 は相対的な意味しかない (例) 温度、西暦
	(比例尺度)	0 は絶対的な意味がある (例) 身長、体重、年齢

## 4 2つのデータの関係性の分析

- ・ ( ) = ある要素と別の要素が互いに影響し合っている状態のこと
  - ・ ( 正の相関 ) = 一方が増えると、もう一方も増える関係
  - ・ ( 負の相関 ) = 一方が増えると、もう一方が減る関係
- ・ ( ) = ある要素が原因となって別の要素が結果として発生する関係のこと  
→ 相関関係があるからと言って、必ず因果関係があるとは限らない!

### 【確認課題】調べよう・考えよう!

(1) 次のグラフのデータから分析してみよう。



【振り返り】 No.22 の実習・学習で学んだこと、気づいたこと、考えたことを 3 行以上書きましょう。