

# 高等学校情報科「情報Ⅰ」 教員研修用教材

## 第4章

## 情報通信ネットワークとデータの活用

## 第4章

# 情報通信ネットワークとデータの活用

### ◆本単元の学習内容

【学習内容の全体像】

#### (4) 情報通信ネットワークとデータの活用

(ア)

情報通信ネットワーク  
の仕組みと役割

- (1) 情報通信ネットワークの仕組みと役割  
サーバ, クライアント, ルータ, 無線 LAN
- (2) 通信プロトコルとデータ通信  
プロトコル, 経路制御, 伝送制御
- (3) 情報セキュリティ  
個人認証, デジタル署名, 暗号化

(イ)

情報システムと  
データの管理

- (1) データの蓄積と管理  
データ, データの蓄積
- (2) データベース  
関係データベース
- (3) 情報システムとそのサービス  
POS, ATM, 情報システム, トレーサビリティ
- (4) データの提供  
オープンデータ

(ウ)

データの  
収集・整理・分析

- (1) データの表現  
尺度水準, 量的データ, 質的データ, データの形式, ビッグデータ
- (2) データの収集と整理  
Web スクレイピング, データクリーニング
- (3) データの分析と評価  
テキストマイニング, 単回帰分析, 可視化, 評価指標

(全体)

情報通信ネットワークの管理, 運用ができ, データベースや Web 上のテキストデータ, オープンデータ等を可視化, 分析する力を育成する。

### 【学習目標】

- 情報通信ネットワークや情報システムにより提供されるサービスを活用する活動を通して情報の科学的な見方・考え方を働かせて、情報通信ネットワークや情報システムの仕組みを理解する。
- データを蓄積，管理，提供する方法，データを収集，整理，分析する方法，情報セキュリティを確保する方法を身に付けるようにし，目的に応じて情報通信ネットワークや情報システムにより提供されるサービスを安全かつ効率的に活用する力やデータを問題の発見・解決に活用する力を身に付ける。
- 情報技術を適切かつ効果的に活用しようとする態度，データを多面的に精査しようとする態度，情報セキュリティなどに配慮して情報社会に主体的に参画しようとする態度を身に付ける。

### 【本単元の取扱い】

- (ア)については，小規模なネットワークを設計する活動を取り入れるものとする。
- (イ)については，自らの情報活用の評価・改善について発表し討議するなどの活動を取り入れるものとする。
- (ウ)については，比較，関連，変化，分類などの目的に応じた分析方法があることも扱うものとする。

### 【中学校までの学習内容との関連】

- 情報通信ネットワークの仕組み，情報システムにおけるデータを通信する技術やデータを蓄積，管理，提供する方法については，中学校技術・家庭科技術分野の内容「D 情報の技術」の(2)のAに示す「情報通信ネットワークの構成と，情報を利用するための基本的な仕組み」についての学習を踏まえて扱う。また，統計的な内容については，中学校数学科の領域である「D データの活用」を踏まえて扱う。

## ■研修内容

### 【研修の目的】

- 小規模な LAN を構成するのに必要な機器やその接続状況を理解できる。
- 有線 LAN と無線 LAN の違いやその接続に際しての注意点について理解できる。
- 情報通信ネットワークにおけるデータの流れをプロトコルを用いて説明できる。

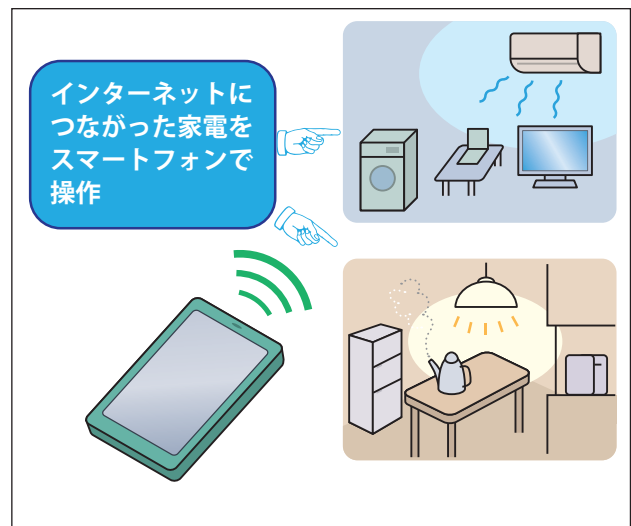
### (1) 身近な LAN について考えてみよう

LAN(Local Area Network) には様々な規模のものがある。家庭やコンビニエンスストア、歯科医院のような小規模な LAN、学校や企業の営業所などの中規模な LAN、企業や都道府県内の学校ネットワークなどの大規模な LAN である。このような LAN の規模とは何をもとに考えたらよいのだろうか。

最も身近な家庭内の LAN や学校のコンピュータ室の LAN について考えてみよう。家庭内の LAN は何のために構成されているのだろうか。いくつかの情報機器がインターネットへの接続を必要としているのでそれをまとめるために LAN を構築していることもあれば、家庭内で録画してあるビデオや保存してある写真、ファイルなどを共有できるように LAN を構築していることなどが考えられるだろう。これらの家庭内の LAN については、既に外部への接続（ルータ等）の部分は確保されている場合もあれば、回線を契約して、必要な接続機器を設置してもらっていたり、携帯電話のキャリアや CATV が提供している機器などを導入していたりと様々な形態をとっていることもある。ルータ等の LAN の内部と外部を分ける機器については、どのような役割を果たしているのか、IP アドレスや簡単なプロトコル、ファイアウォールなどの関係などととも生徒に理解させることが必要である。学校のコンピュータ室などでは、コンピュータ室の LAN が学校内の LAN の一部である場合や、コンピュータ室の LAN が学校内の LAN から独立している場合の違いにより、その仕組みや活用方法も異なってくることも理解が必要である。

さて、家庭内の LAN に話を戻そう。家庭内の LAN に接続している機器にはどのような機器があるだろうか。単なる情報機器だけではないだろう。パソコン、タブレット、スマートフォン、プリンタなどの情報機器だけでなく、テレビ、録画用のハードディスクレコーダ、電話、電子レンジなど、調べてみると、さまざまな機器があげられることだろう。またこれらを接続するためにハブや無線のアクセスポイントなどの機器もあるかもしれない。1つの家庭の LAN に何台の機器が接続されているか。有線で接続されているか、無線で接続されているかを調べる実習などを、実際の生徒に課すことにより、家庭電化製品とインターネットとの関わり、IoT (Internet of Things) への理解を深めることができる。

同様に学校のコンピュータ室やコンビニエンスストア、レンタル CD ショップ、歯科医院などの小規模な店舗についても考えてみる。学校のコンピュータ室に関しては、接続機器の台数は多いと考えるが、小規模な店舗などでは、家庭と同程度かまたはそれより少し多い機器が接続されているだけである。店舗や学校のコンピュータ室のように数十台までの接続機器が参加している LAN をここでは、小規模な LAN と考えてみる。



図表 1 IoT の一例

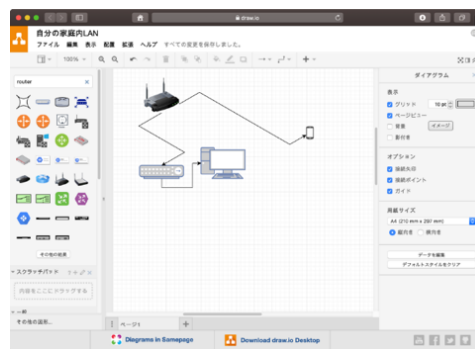
### <演習 1>

歯科医院(または小規模な店舗)を想定し、それをもとに小規模な LAN の構成に必要なものを列挙してみましょう。店舗内の機器の配置図も描いてみましょう。

この演習については、紙のワークシートに手書きで描いてもよいが、コンピュータを利用して描いてみることをお勧めする。一般的なワープロソフトウェアを利用してよいが、draw.ioなどの専用のソフトウェアを使ってみるとよいだろう。



図表 2 draw.io  
出典：draw.io スクリーンショット



図表 3 家庭用 LAN の作図例  
出典：draw.io スクリーンショット

draw.io は、このようなネットワークの図やフローチャートなどの流れ図などを描くことのできる無償のソフトウェアである。ダウンロードして使うバージョンもあるが、コンピュータへのフリーウェアなどのインストールの危険性を危惧するならば、Web 版の draw.io を使用することをお勧めする。

URL : <https://www.draw.io/>

Web ブラウザまたはアプリケーションを開くと最初に図表 2 のような画面が表示されるので、言語を設定したのちに「新規ファイルを作成する」を選択して、始めることができる。

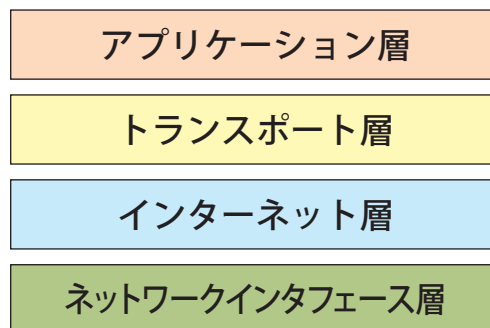
残念ながらパーツの選択に関して、検索は英語のみしかできないが、楽しみながら家庭内 LAN の図が書けると考える。

## (2) 有線 LAN と無線 LAN の違い

TCP/IP の 4 階層モデルについて説明する。インターネットの通信において、通信する際の規定を定めたものを通信プロトコル (略してプロトコルということもある) という。このプロトコルがあることで、相互に通信を行うことが可能になっている。TCP,IPv4,IPv6,DHCP,HTTP,SMTP,POP など様々なプロトコルがインターネットの通信で用いられるが、これらは、4 層のレイヤに分けて考えることができる。インターネットでアプリケーション同士が通信する際は、上位のアプリケーション層から下位の層に移動することで、通信手順が付加され、ネットワークインタフェース層で物理的な通信を行っていく。

さて、有線 LAN と無線 LAN の違いとはどこにあるのだろうか。

外見上は、有線が Ethernet (イーサネット) (IEEE802.3 (アイトリプルイーハチマルニテンサン)) に従った規格の通信であり、無線は LAN ケーブルなしで Ethernet の通信技術を実現したものである。物理的な違いであり、実際には通信上のネットワークインタフェースが Ethernet の電気信号であることと電波を利用することの違いである。急速に無線 LAN は普及しており、2020 年の東京オリンピック等の影響もあり、公衆 Wi-Fi スポットの増加など、これからも無線 LAN は広がっていく傾向になるだろう。



図表 4 4 層のレイヤ

### <演習 2>

無線 LAN のメリットやデメリットに関して、下のような表にまとめています。これについて補足するとともに、有線 LAN についても同様にまとめてみましょう。

	メリット	デメリット
無線 LAN	<ul style="list-style-type: none"> <li>・設置工事の必要がほとんどない</li> <li>・ハブや LAN ケーブルなどが不要</li> <li>・LAN ケーブルが邪魔にならない</li> </ul>	<ul style="list-style-type: none"> <li>・一般的に通信速度が遅い</li> <li>・接続機器の特定が困難</li> <li>・電波の干渉を受けやすい</li> <li>・簡単に通信内容を傍受できる</li> </ul>
有線 LAN		

図表 5 無線 LAN と有線 LAN のメリット・デメリット

無線 LAN, 現在では IEEE802.11 で規格化され, 通信品質や速度の問題に関しては, あまり問題はなくなってきている。無線ルータに書いてある IEEE802.11 の種類と対応については理解しているだろうか。IEEE802.11 について, 製品に記載されている規格, 周波数帯, 最大速度などについて, 生徒に調べさせる実習なども行ってもよいだろう。

さて, 無線 LAN には, 暗号化の方式も複数存在している。また, 公衆無線 LAN の中には, 暗号化されていないアクセスポイントも多く存在している。スマートフォンのモバイルデータ通信容量を節約するために, ファーストフードチェーンやコンビニエンスストア, 携帯電話キャリア等の設置している公衆無線アクセスポイントを活用している生徒も多い。また, それ以外でも公共交通機関, 自治体等が設置しているアクセスポイントも多く見つかる。このような公衆無線 LAN で, どのような暗号化方式が使われているのか, 暗号化が使われていないまたは安全性の低い暗号化方式が使われているアクセスポイントではどのような点に注意すればよいのかについて意識させることも大切な情報スキルであると言える。

### <演習 3>

身近にある公衆無線 LAN でどのような暗号化が用いられているのか調べてみましょう。

番号	公衆無線 (SSID 等)	暗号化 (AES 等)	認証 (Web 認証等)	時間制限
1				
2				
3				
4				
5				

図表 6 暗号化

さて, 暗号化方式については, OS や通信機器のメーカーによって様々な呼び名を用いているので整理してみる。大まかに整理すると, 次の表のように認証方式と, 暗号化アルゴリズムのペアで用いられていることが多い。

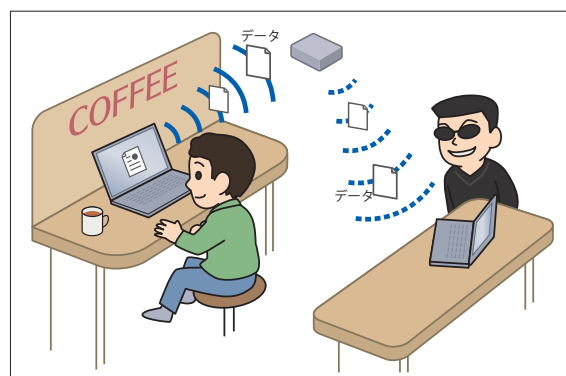
	認証方式 (セキュリティの種類)	暗号化方式	暗号化アルゴリズム
1	WEP	WEP	RC4
2	WPA	TKIP	RC4
3	WPA2	CCMP	AES
4	WPA3	SAE	CNSA

図表 7 認証方式と暗号化アルゴリズム

表を見てわかるように製品によって, 認証方式を示しているのか, 暗号化アルゴリズムを示しているのかがわからない部分もあるので, 整理しておく必要がある。有線 LAN では物理的なケーブルを用いてデータを転送するのに対して, 無線 LAN では, 空気中を電波でデータを伝送するために, 傍受することが容易であることがデメリットである。

共通鍵暗号ならば, 暗号化, 復号の両方で同じキーを使うので, この方式を使う無線 LAN は, キーがわかっしまえば, 安全な通信はできないことになる。

このうち, WEP は古くから使われている方法ではあるが, 現在ではフリーウェアを用いて暗号化キーが数秒で解読されてしまうほど, 脆弱な暗号であることが知られ, 現在ではほとんど使われていない。そのため安全性を高めるために WPA-TKIP や更にセキュリティを高めた WPA2-CCMP(AES) が用いられていたが, 2017 年 10 月 16 日に『Wi-Fi Protected Access 2 (WPA2)』に存在する脆弱性が複数発見された」(<https://www.krackattacks.com>) との報道により,



図表 8 公衆無線 LAN の危険性 (データ傍受)

全世界は震撼した。これによって、新たな規格である WPA3 が検討されてきている。

家庭内 LAN では、Wi-Fi は今や欠かせない通信技術だが、このように脆弱な部分や危険性についても理解しておく必要もある。公衆無線 LAN では、いまでも WEP という、脆弱な技術を使用している場所もあるし、そもそも暗号化方式を設定していない無線アクセスポイントも多くある。生徒自身が家庭内で無線アクセスポイントを設定する権限がなかったとしても、自身の PC で Microsoft Windows などの機能で無線ホットスポットを設定した場合、不正侵入の入り口を作ってしまう危険性もある。これにより、内部の情報機器への侵入が可能ということも理解させたい。そのために公衆無線 LAN でしてよいことと、問題がある危険性を持つことに分けることが必要になる。また、安全が確保されていない無線 LAN では、VLAN を設定するなどの知識も生徒に理解させることも必要になる。安全性が確保されていない無線 LAN には、できるだけ接続しないようにする。

#### <演習 4>

安全性の低い公衆無線 LAN で、しても大丈夫なことと、危険性のあることについて、話し合ってみましょう。また、なぜ大丈夫なのか、なぜ危険なのかを明確にしてみましょう。

実際の小規模な無線 LAN の構築に関しては、学習 19 で触れる。その際にも無線ルータやアクセスポイント、暗号化方式などの知識が必要になる。また、家庭向けの無線ルータには、ゲストモードなどの外部からオープンで接続できるアクセスポイントを設定できるものもある。これらの接続の際に留意することに関しても、授業の中では、生徒の理解を深めるために議論させることが大切となる。不用意に購入してきたままの設定で使っている家庭や個人も多いことが、トラブルの原因を作っていることもある。理解した上で、きちんと設定ができるように、生徒の知識と技能を高めることが大切である。

### (3) 情報通信ネットワークのプロトコル

情報通信ネットワークのプロトコルに関しては、現行の学習指導要領でも触れられており、多くの教科書でも扱われているが、形式的に覚えさせる内容ではなく、実際に何が行われているのか、それぞれのレイヤが何を意味しているのかについて、いくつかの例を理解させることが必要になる。お互い顔が見えない場所でトランシーバーや糸電話で通信を始めるとき、どのような事前の約束事が必要になるのか、それぞれの約束事やモノが、実際の情報通信ネットワークでは、何に対応しているのか。また、複数人での一斉通信を行う Skype やテレビ電話などを行うときに、お互いにどのようなことに配慮すればよいのか、などを考えることからプロトコルの一部を理解する学習活動が考えられる。

#### <演習 5>

生徒 5 人が定期考査前に、テレビ会議を使って勉強をする場面を想定します。事前に約束しておく内容や話し合う際に留意すべき内容について、どのようなものが考えられるか、整理してみましょう。

#### <参考文献・参考サイト>

- ・「KRACK Attacks Breaking WPA2」<https://www.krackattacks.com/>
- ・「draw.io」<https://www.draw.io/>

## ■ 学習活動と展開

### 【学習活動の目的】

- 情報通信ネットワークのプロトコルの種類とその役割について、理解する。
- インターネットのプロトコルの種類と基本的なプロトコルの役割や仕組みについて理解する。
- 無線 LAN の仕組みとそのセキュリティについて、暗号化も含めて理解する。
- 電子メールや Web の仕組みについて、そのプロトコルとともに理解する。
- プロトコルとそれを悪用したサイバー犯罪に関して理解する。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	情報通信ネットワークのプロトコルの必要性を理解しよう。	プロトコルがなかったら、データを転送する際にどのようなことが起こるのか、話し合う。
展開 2	公衆無線 LAN に接続することは危険ではないのだろうか。どのようなことに注意して接続すればよいのか考えてみよう。	身近にある公衆無線 LAN についてその種類と暗号化方式について調べる。また、簡単な接続できることと危険性のトレードオフについて考える。
展開 3	安全な接続は、不便な接続なのか、考えてみよう。	情報通信ネットワークに安全に接続するには、どのようなことを留意すればよいのかについて考えさせる。また、なぜ必ずしも安全でないネットワークが存在するのかについて、様々な場面を想定して考えさせる。

### 展開 1

問 い	情報通信ネットワークのプロトコルの必要性を理解しよう。
学習活動	<ul style="list-style-type: none"> <li>• 糸電話やトランシーバー、3人以上の人で会話するテレビ電話などの具体例をもとに、プロトコルがない場合、どのようなトラブルが生じる可能性があるのか、意見を共有する。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>• 活動につまずいている生徒に、必要に応じてヒントを与え、生徒自らがプロトコルの必要性に気づくよう促す。</li> </ul>





## 展開 2

問 い	公衆無線 LAN に接続することは危険ではないのだろうか。どのようなことに注意して接続すればよいのか考えてみよう。
学習活動	<ul style="list-style-type: none"><li>無線 LAN の規格や暗号化方式について理解を深める。</li><li>手軽に接続できることとセキュリティを保つことがトレードオフの関係にあることを理解する。</li><li>グループでの話し合いの中で、注意すべきことを共有する。</li></ul>
指導上の留意点	<ul style="list-style-type: none"><li>具体的な例を元に様々な意見を採り上げ、危険性について考えることができるように促す。</li></ul>



## 展開 3

問 い	安全な接続は、不便な接続なのか、考えてみよう。
学習活動	<ul style="list-style-type: none"><li>セキュリティを保ちながら使いやすい無線 LAN 環境を作るには、利用の目的などを明確にすることが重要であることを家庭内 LAN と学校 LAN の比較などから理解する。</li></ul>
指導上の留意点	<ul style="list-style-type: none"><li>身近な家庭内 LAN、学校 LAN、公衆無線アクセスポイントを比較することによって、それぞれを利用する目的の違いや留意点について考えられるように促す。</li></ul>



## まとめ

まとめ	<ul style="list-style-type: none"><li>今後、情報通信ネットワークを利用する際に無線 LAN のセキュリティについて意識して、利用目的などを考慮して活用すること必要があることを理解する。</li></ul>
-----	---

## ■研修内容

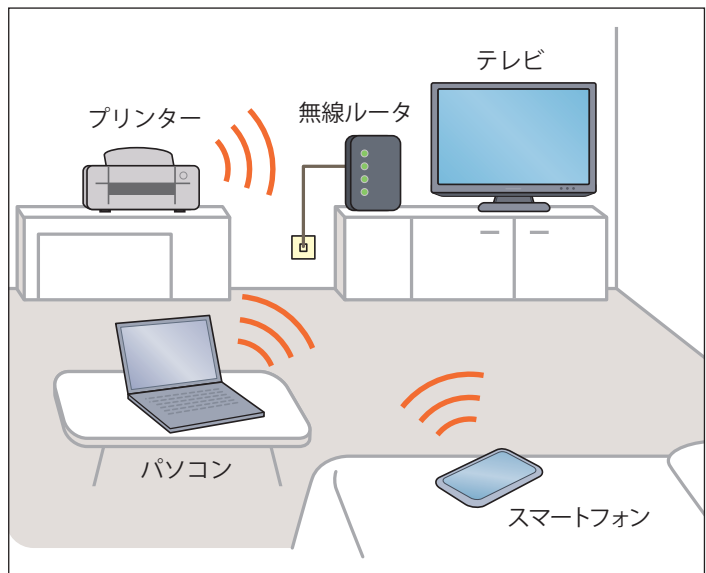
### 【研修の目的】

- 小規模な LAN を構成するのに必要な機器やその接続方法を理解できる。
- 身近な情報機器を小規模な LAN に参加させる方法が理解できる。
- LAN を構築する際に必要なセキュリティについて理解できる。
- 簡単なネットワークのトラブルシューティングができる。

### (1) 小規模な LAN の構築と情報機器

家庭内 LAN のような小規模な LAN に必要な機器に関しては、学習 18 でも扱ったが、ここでは、実際に必要な機器を用いて LAN を設計し、構築していくことを考えよう。

まず、小規模な LAN を設計するには、LAN の利用目的と必要な機器の選定が必要である。家庭内 LAN ならば、接続する機器は既に定まっているだろうし、そこに無線 LAN をもとにした小規模な LAN 構築に必要な機器を考えることから始まる。情報機器が IEEE802.11 のどの規格に対応しているのか、インターネット接続に関して、どのようなプロトコルが必要なのか、LAN を分割する必要があるのか。無線 LAN のアクセスポイントをいくつ設置すればよいのか。ルータとアクセスポイントやハブとの接続の環境はあるのか、などの様々な問題が考えられる。細かい仕様に関しては、調べることが可能であるので、物理的な接続に関して家庭内 LAN の設計を考えるとよいだろう。また、各機器の設定に関しては、「(2) 情報機器を LAN に参加させる方法」で考えていこう。LAN の設計に関しては、学習 18 で紹介した draw.io(アプリケーションまたは Web アプリ (<https://www.draw.io/>)) を用いて、LAN の構成図を書いて考えるとよい。



図表 1 家庭内 LAN

### <演習 1>

4 人家族の家庭内に 2 台の PC、2 台のタブレット、テレビに接続するスティック型 PC1 台、テレビ録画用のハードディスク、スマートフォン 4 台があります。これらは構築する無線 LAN 等接続する情報機器です。これに無線ルータ 1 台、無線アクセスポイント数台を設置して、家庭内の無線 LAN 等を構築した図を描きましょう。ただし、外部のインターネット回線に関しては、無線ルータに有線で接続するだけで利用可能であり、部屋は 3LDK で構成されており、リビングルームでも無線 LAN 等が利用可能であるものとします。

※ここでは仮想的に情報機器を設定し、その構成図を描く演習としたが、実際に情報機器を数台用いて、無線 LAN 等の構築の実習が行えることが望ましい。その際に、無線ルータの設定や IP アドレスの設定、その他の設定などについての知識も必要になる。

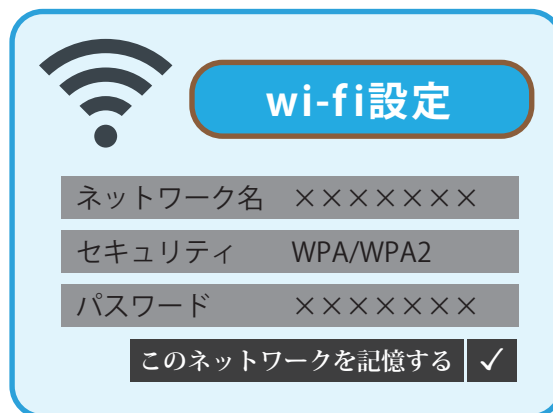
### (2) 情報機器を LAN に参加させる方法

自分で構築した小規模な無線 LAN に機器をどのようにしたら参加させることができるのだろうか。これは、既にある無線 LAN にスマートフォンなどを接続した経験がある生徒ならば容易に理解できるだろう。家庭内などの接続する人が特定できている場所での接続に関しても、外部からの不正な侵入も考慮し、SSID を公開しないこと

や暗号化キーを分かりづらいものにしておくこと、暗号化方式を容易に解読されないものにしておくことが必要である。そのため、SSID、暗号化キー、暗号化方式などの情報が無線 LAN に参加するのに必要になる。家庭内の情報機器によっては、IP アドレス等を設定することが難しい機器などもあることから、DHCP の IP アドレスの配布や WPS などの簡易的に接続する方式などを利用することも多くあるだろう。

しかし、これが学校や企業内 LAN、さらには公衆無線 LAN に自分の情報機器を参加させる場合には状況が異なってくる。特定多数あるいは不特定多数の人が無線 LAN を利用するからである。そこでは、LAN への参加の容易さとセキュリティとのトレードオフがある。

学校や企業内 LAN に自分の機器を参加させる機会は少ないかもしれないが、デジタル教科書などの活用が始まるこれからの学校では、学校内に無線 LAN の環境を整備し、生徒個人の情報機器を LAN に参加させる必要が生じることが考えられる。一方で、学校の生徒は容易に接続できるが、外部の人に容易に接続できないような工夫が必要である。また、コンビニエンスストアやファーストフードチェーン、公共交通機関、地方自治体等が提供する公衆無線 LAN は接続の容易さと引き替えにセキュリティに関しては、低く設定されている場合もある。これらの公衆無線 LAN に関しては、大手の通信キャリアが回線を提供している場合なども多いが、その利用やセキュリティに関しては利用者個人の責任となる場合もあるので利用には注意が必要である。学習 18 の演習では、このような公衆無線 LAN について、調べる実習を用意している。



図表 2 ネットワークの記憶

## <演習 2>

公衆無線 LAN に参加する場合のセキュリティや個人の責任に関しては、どこに明記されているのか調べてみましょう。また、電子メールのチェックや Web サイトの認証などにおいて、暗号化がなされているのか調べてみましょう。

### (3) 無線 LAN の構築に関するセキュリティ

LAN を構築する際に必要なセキュリティについて理解できる。

前節では、物理的な LAN 接続について考えたが、本節では設定について考えてみよう。まず、無線ルータや無線のアクセスポイントをどのような設定にしているのか知ることが必要である。

無線の接続に関しては、IEEE802.11 の規格、暗号化方式、SSID の設定、DHCP の設定、フィルタの設定など、様々な設定内容が用意されている。これらの内容を理解し、適切な設定ができることが大切である。

無線ルータを購入した設定のまま接続し、利用している家庭について安全面で不安な部分が多く考えられる。それに対して、ある程度の設定ができるスキルを身に付けることが、無線 LAN が主流となりつつある現在でインターネットを活用する生徒に必要なであろう。

無線ルータや無線アクセスポイントの暗号化方式に関しても学習 18 で説明したように脆弱性の見つかったものなども多くある。どのような暗号化方式を利用すればよいのか。SSID は外部から参照可能なものにした方がよいのか、見えないようにすべきなのか、その危険性についてなど話し合いをすることも必要である。暗号化キーの管理に関してもパスワードの管理と同様にわかりやすいものにすれば、悪意のある者からも解読されやすくなり、セキュリティが低下する、というトレードオフの部分をどのように解消するのかも問題となる。また、情報機器に IP アドレスをリースする DHCP(Dynamic Host Configuration Protocol) に関しても、設定する方がよいのかどうかで議論の分かれるところである。家庭内と学校や企業内では、セキュリティの基準が異なるので、設定が異なるものになるのは当然のことである。

これらの演習に関しては、先に挙げたようなルータの設定画面などを見ながら、実際に設定してみると理解が深まるだろう。



図表 3 無線 LAN の傍受

### <演習 3>

- (1) 家庭に小学校高学年の子どもがいて、夜 12 時以降には接続できないようにしたいです。与えられた無線ルータのフィルタ設定で、どのように設定すればよいか考えてみましょう。
- (2) 親戚の子どもが夏休みに 10 日間、自宅に泊まることになりました。自宅の Wi-Fi 環境を使わせたいですが、通常使用している暗号化キーは教えたくありません。どのような接続が考えられるでしょうか。

自宅で友人や外部の人たちに Wi-Fi 環境を提供したい場合などにどのように対応できるかとの問題である。このような場合は少ないかもしれないが、自分がカフェなどの店舗のオーナーであったとき、どのように業務用の無線 LAN とお客様へのサービスとしての無線 LAN 環境についてセキュリティを保ったまま維持できるか、という問題と同様である。このような仮想の設定などで対応を考えることで、実際の LAN の構築に対応するスキルを身に付けることができる。インターネット回線が 2 つある場合、ネットワークをどのように分割すればよいか。様々な方法を考えることが可能である。

### <演習 4>

現在、自宅の無線 LAN に接続されている機器の MAC アドレスや割り当ててある IP アドレスの一覧を調べる方法はあるでしょうか。実際にこの実習環境で接続されている機器を特定し、その情報機器に割り当てられている IP アドレス等を調べてみましょう。

## (4) ネットワークのトラブル対応

LAN においては、様々なトラブルが生じることがある。

- ・インターネットに接続できない。
- ・無線 LAN のアクセスポイントに接続できない。
- ・電子メールを送信できない。
- ・必要なアプリをダウンロードできない。
- ・家庭内の HDD やサーバに接続できない。
- ・家庭内のネットワークプリンタに出力できない。

など、様々な問題が起こることがある。

これらの問題には、どのような対応をすればよいだろうか。すぐにインターネットの回線業者に電話し、修理を依頼するのだろうか。また、ルータや PC を新規に購入するのだろうか。

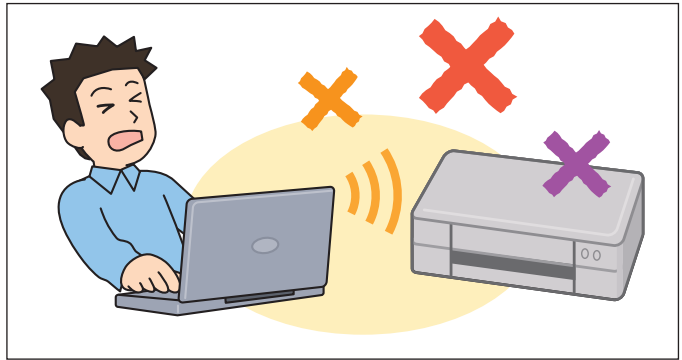
まずは、トラブルの原因を特定し、それが自分で復旧できる内容ならば対処し、自分で対応できないものであれば、業者に修理を依頼する、または新規に購入するという判断になるだろう。それでは、どのようにトラブルの原因を特定すればよいのだろうか。

例えば、ネットワークに接続されたプリンタに出力できない場合には、以下のような問題が考えられる。

- ① プリンタの電源が入っていない。
- ② プリンタが紙切れ、インク切れになっている。
- ③ プリンタにネットワークの設定がされていない。
- ④ プリンタにネットワークの接続がされていない。
- ⑤ プリンタに接続しているネットワークの機器に不具合がある。
- ⑥ 印刷の命令を送っていない。
- ⑦ 印刷するためのアプリが適正に設定されていない。
- ⑧ 自分の情報機器が LAN に接続されていない。
- ⑨ 自分の情報機器の LAN の接続の設定が誤っている。
- ⑩ 自分の情報機器のプリンタの設定がされていない、または誤っている。

などである。次にこれらを点検する順序に並べてみる。自分の情報機器からプリンタへの順序でも、自分が点検しやすい機器から点検していくという順序でもよいと考える。大切なのは、それぞれの問題の点検項目で何が特定できるかということである。例えば、④の問題を点検するには、プリンタの LAN ポートにケーブルが接続さ

れているのか、または無線 LAN であれば、適当なアクセスポイントに接続できているかを点検すればよいだろう。それによって、プリンタと LAN との物理的な接続に障害があるのが特定できる。有線の場合、プリンタ側の LAN 接続を示す LED がついているかどうか目安となる。⑤とも関係しているが、プリンタに確かにケーブルは接続されているが、LED がついていないときなどは、ケーブルを交換してみる、同じ HUB のプリンタが接続されていた同じポートに別な情報機器を接続して試してみる、同じ HUB の別なポートにプリンタを接続してみる、などを行うことで、プリンタ側の LAN 接続装置の異常や HUB のポートの異常、ケーブルの不良などを調べることもできる。つまり、どの部分を調べるかによって、障害の切り分けが可能なのである。これらを効率よく行い、障害を特定していくためには、どのような順序で点検すればよいかを考えればよいのである。また、ケーブルの不良ならば別なケーブルに交換すればよいが、LAN 接続装置の故障ということになれば、プリンタを修理に出すか、新規に購入するという対応になる。



図表 4 プリンタの接続トラブル

また、PC やスマートフォンからプリンタに接続している場合、そのプリンタが PC などの LAN に接続されている機器などから、見ることができるかどうか調べる技術に関しても学んでおきたい。プリンタドライバなどから、LAN 上のプリンタの状況をどこまで監視できるのか、調べてみることもこの演習の応用として考えることができる。

<演習 5>

自宅の Wi-Fi に繋がっているスマートフォンからインターネットに接続ができなくなりました。昨日までは正常に接続できており、直前に OS のアップデートなどはしていません。どのような問題が考えられるでしょうか。問題を列挙し、それらを点検する順序に並べてみましょう。またその点検を行うことによって、どの部分の障害を切り分けることができるのかを考えてみましょう。

	問題	点検方法	障害の箇所
1			
2			
3			
4			
5			

図表 5 ネットワークのトラブル対応

情報通信ネットワークにおいて、LAN 上のトラブルの障害の切り分けは、主に利用者が行うとよいだろう。トラブルへの対応を外部へ依頼するか、自分で対応できるかを適正に切り分ける能力を育成することによって、外部に修理等を依頼する際にも適切な説明ができるような力を育成することができると考えている。

## ■ 学習活動と展開

### 【学習活動の目的】

- ルータ、ハブ、コンピュータなどから、どのように LAN を構築できるか理解する。
- 小規模な無線 LAN を構築する方法について理解する。
- ネットワークが接続できない、ネットワークに接続されている機器に接続しないなどのトラブルにどのように対応するのか理解する。
- 小規模な LAN が家庭も含めどのようなところで活用されているのかを理解する。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	家庭でインターネットに接続できる機器には何があるでしょうか。またそれらの機器を分類し、構成図を描いてみよう。	家庭でインターネットに接続できる機器は何台あるのか、それらを接続する際に有線と無線の区別、家電と情報機器の区別、サーバの役割を果たしている機器などについて考える。家庭内 LAN の構成図を描いてみよう。
展開 2	家庭で無線 LAN を構築するには、注意点をまとめてみよう。	DHCP や暗号化キーについてどのような設定をすればよいか。また、フィルタなどの仕組みについて調べる。
展開 3	家庭内 LAN に関するトラブルについて考えてみよう。	家庭内 LAN を利用する際に生じるトラブルに対して、どこに障害があるのか、特定できる方法を考えよう。

展開 1	
問 い	家庭でインターネットに接続できる機器には何があるでしょうか。またそれらの機器を分類し、構成図を描いてみよう。
学習活動	• 家庭内の LAN の構築を考えるためにインターネットに接続している機器を分類して、その構成図を描くことで、理解を深める。
指導上の留意点	• ネットワークという、日常的に利用はしているものの、視覚的にはあまり確認したことがないものを取り扱うため、理解に結びつくように説明する。



展開 2	
問 い	家庭で無線 LAN を構築するには、注意点をまとめてみよう。
学習活動	<ul style="list-style-type: none"> <li>・家庭内の無線 LAN を構築するためには、どのような危険性があるのか、話し合いの中で、共有し、構築できるようにする。</li> <li>・無線 LAN を構築するために必要な知識や技術を身に付ける。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>・技術的に難しい内容については、理解を深められるように丁寧に説明する。</li> </ul>



展開 3	
問 い	家庭内 LAN に関するトラブルについて考えてみよう。
学習活動	<ul style="list-style-type: none"> <li>・家庭内や学校などにおいて生じるネットワークのトラブルについて、自ら対応して、障害の箇所を切り分けたり、正しい点検の方法を理解させる。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>・日々なんとなく実行していることを明確化することにより、無駄なく複雑なトラブルを解決できる道筋を明確にできるように促す。</li> </ul>



まとめ	
まとめ	<ul style="list-style-type: none"> <li>・生徒自らが身近なネットワークの単なる利用者ではなく、その管理と運用ができるようになることの必要性を理解させる。</li> </ul>

## ■研修内容

### 【研修の目的】

- 情報システムにより提供されているサービスについて理解する。
- オープンデータの重要性とその活用について理解する。
- 情報システムにより活用しやすいデータの形式について理解する。
- 公的なオープンデータを活用するための方法について理解する。
- 公的なオープンデータを活用して、社会的な問題を発見する授業ができるようになる。

### (1) 情報システムが提供するサービス

私たちの生活や様々な産業の中で、様々な情報システムが活用され利便性が向上している。オンラインショップのシステムや電子決済システムなどにより便利に買い物ができるようになってきている。情報システムが提供するシステムは行政、医療、教育など幅広く構築され、活用されている。主な情報システムには、図表1のようなものがある。

これらの情報システムは、情報通信ネットワークにより様々なデータが送受信されて運用されている。情報システムにより蓄積された膨大な量のデータ（ビッグデータ）を課題解決に活用することが近年重要視されている。個人情報を含むデータや企業活動の根幹にかかわるデータは、漏洩することがないように情報セキュリティを確保する必要がある。一方、公益性が高いデータについてはオープンデータとして公開される動きも広がっている。

分野	情報システム	活用の例
商業	POS システム (Point Of Sales System)	店舗経営の効率化・売上分析・在庫管理
金融	電子決済システム	鉄道運賃支払, 店頭レジでの支払い
通信	SNS (ソーシャル・ネットワーキング・サービス)	利用者同士の交流, 情報発信
交通	高度道路交通システム (ITS)	渋滞緩和, バスの走行位置提供
気象	緊急地震速報システム	早期避難, 工場の機械制御
教育	e ラーニングシステム	教材提供, 学習履歴に基づく指導
行政	住民基本台帳システム	行政手続きの簡略化, 住民票写し広域交付

図表1 情報システムが提供するサービスの例

### (2) オープンデータの重要性とその活用

国や地方公共団体及び事業者が保有するデータには、誰もがインターネットを通じて得られるよう公開されるものが増えつつある。このうち、次のような項目に該当する形で公開されたデータはオープンデータと呼ばれている。

#### ■オープンデータの定義■

- ① 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
- ② 機械判読に適したもの
- ③ 無償で利用できるもの

出典：「オープンデータ基本指針」（総務省）  
<https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/kihonsisin.pdf>



このうち「機械判読に適したもの」として、「コンピュータプログラムが自動的にデータを加工、編集できること」ができるようなものも含め、オープンデータを普及させる取組が官民をあげて推進されている（官民データ活用推進基本法（平成 28 年制定・施行））。このようなオープンデータが公開されることにより、次のような効果が期待されている。

### ■ オープンデータとしてデータが公開される意義・効果 ■

- ① 国民参加・官民共同の推進を通じた諸課題の解決、経済活性化  
多様なサービスの提供・公共サービスの提供・改善、諸問題の解決への貢献
- ② 行政の高度化・効率化  
情報を根拠とした政策・施策の企画・立案、効果的な行政の推進
- ③ 透明性・信頼の向上  
政策等に関する十分な分析・判断が可能、行政の透明性・信頼性の向上

「オープンデータ基本指針」（総務省）

(<https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/kihonsisin.pdf>) を加工して作成

### (3) データの蓄積・管理・提供する方法

データを活用するためには、容易にデータを処理する必要がある。特に災害発生時のように緊急を要する場合には、迅速性が求められる。紙の資料を読み取った画像だけの PDF ファイルを活用するには、改めてデータを入力することが必要となる。また、表計算ソフトウェアに入力されたデータであっても、印刷するための書式では、データを活用するために整形することが必要となる。一方、プログラムを用いてデータを二次利用する場合に、機械判読が可能となっていれば、さまざまなデータを組み合わせて新たな情報を生み出すことができる。

オープンデータが十分に活用されるよう、総務省による研修資料では、図表 2 のようにファイル形式を掲載し、利用しやすいデータの公開を推進している。

ファイルの種類	ファイル形式	
統計情報、施設一覧など表形式となっている情報、予算・決算情報	表形式	Excel 形式 CSV 形式など
報告書など、文章による情報や、図形、画像等が混在している情報	文書形式	PDF 形式 Word 形式など
主に地図と組合せて利用する情報で、線や面などのデータを含むもの	地理空間情報形式	※ shp 形式など

※ shp 形式＝シェープファイル形式（図形情報や属性情報を持つ複数のファイルから構成される）

図表 2 オープンデータ担当者向け研修資料に掲載されたファイル形式

出典：「オープンデータ化支援研修～第 2 部 オープンデータを公開するための手順」（総務省）

([https://www.opendata-training.org/wp-content/uploads/material/2\\_support/2\\_new/shien\\_opendata\\_tejun.pdf](https://www.opendata-training.org/wp-content/uploads/material/2_support/2_new/shien_opendata_tejun.pdf))

### (4) 国や地方公共団体が提供するオープンデータ

国や地方公共団体では、様々なデータをオープンデータとして公開している。国により、人口や世帯、経済、産業、安全、環境、教育など省庁が調査したデータが公開されている。また、気象庁が観測した気象データ、国立環境研究所による大気汚染や騒音などの観測データなど、専門的に観測しているデータも公開されている。

地方公共団体では、取り組みにばらつきはあるが、市町村の人口だけでなく、町丁字ごとの人口や避難場所、地方公共団体が設置した AED の設置場所、フリースポットの場所、市内を走るバスに関するデータなど生活に密着したデータを公開している地方公共団体もある。

### (5) GIS を用いたデータの可視化と問題発見

データの活用方法として、最大値、最小値、中央値、四分位数、最頻値、平均値、分散、標準偏差といった基本統計量を求め、各統計指標の状況を整理したり、相関分析や回帰分析等を通して問題発見や問題解決に役立てることが考えられる。そのような統計的な分析だけでなく、データを可視化することにより、問題発見や問題解決ができる。

ここでは、地理情報データを GIS (Geographic Information System) を用いて可視化することを取り上げる。GIS には、無料の MANDARA10、Quantum GIS や有料の ArcGIS などがあり、防災等の地域分析が可能となる機能が提供されている。ここでは、政府統計の総合窓口 (e-Stat) の統計 GIS を用いて、データの可視化と問題発見を行う。

#### <演習 1>

埼玉県川越市に AED が十分に設置されているかどうかを調べてみましょう。

オープンデータである政府統計ポータルサイトから町丁・字等別人口密度のデータと埼玉県オープンデータポータルサイトから川越市の AED 設置場所データを使って、2つのデータを重ねることで問題を発見してみましょう。

アカウントの作成	・地図で見る統計 (jSTAT MAP) のアカウントを作成する。
データの検索・収集 データの整形・加工	・川越市の AED 設置場所データを探しダウンロードする。 ・jSTAT MAP の形式に合わせてデータを整形・加工する。
データの読み込み 地図の描画	・jSTAT MAP を起動して川越市の AED 設置場所を描画する。 ・同様に川越市の町丁・字等別人口密度を描画する。
地図の分析	・描画された地図から、人口に対して AED が少ない場所を見つける。

図表 3 演習の手順

### [手順 1] アカウントの作成

地図で見る統計 (jSTAT MAP) は、ログインしなくてもある程度の機能は使える。しかし、今回は、オープンデータから川越市の公共 AED 設置場所データを jSTAT MAP にインポートして表示する。そのためには、アカウントを作成してログインする必要がある (図表 4)。

アカウントを作成するには、はじめにメールアドレスを仮登録する。仮登録の直後に、本登録用の案内が書かれたメールが自動返送されるので、メールの指示に従って操作することによりアカウントが作成できる。

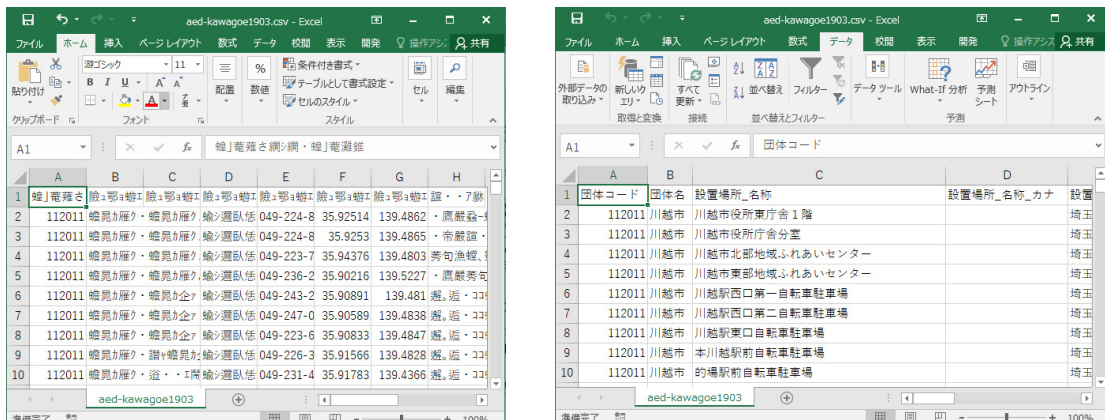


図表 4 政府統計の総合窓口 (e-Stat) の画面  
出典：政府統計の総合窓口 (e-Stat) (<https://www.e-stat.go.jp/>)

### [手順 2] データ検索・収集, データの整形・加工

川越市の AED 設置場所データを埼玉県オープンデータポータルサイト (<https://opendata.pref.saitama.lg.jp>) からダウンロードする。ここでは、緯度・経度が掲載された CSV データを取得している。(各サイトによっては取得方法が異なるため手順は省略する)

次に CSV データを jSTAT MAP に読み込めるよう、Excel を使って整形する。CSV データを開く際に文字コードの影響により、文字化けする場合がある。この場合には、メニューバーで [データ] に切り替え、「外部データの取り込み」の [テキストファイル] のボタンを使って開くことにより、文字コードを指定して開くことができる (図表 5, Excel 2016 の場合)。CSV データを開いたら、最初の列には設置施設名、末尾の 2 列には経度・緯度となるように、列の順序を入れ替える (図表 6)。図表 6 では、不要な列を非表示にしているが、削除してもよい。データの整形が終わったら、保存する。



図表 5 AED 設置場所データを開いた時の文字化けと対処法

出典：【川越市】AED 設置場所 (公共施設), <https://opendata.pref.saitama.lg.jp/data/dataset/aed>  
川越市保健医療部保健医療推進課作成 CCBY2.1jp <https://creativecommons.org/licenses/by/2.1/jp/>

設置場所_名称	設置場所_住所	設置場所_経度	設置場所_緯度
川越市役所東庁舎1階	埼玉県川越市元町1-3-1	139.4861841	35.92513661
川越市役所庁舎分室	埼玉県川越市元町1-3-1	139.4865436	35.9252954
川越市北部地域ふれあいセンター	埼玉県川越市山田1578-1	139.4803489	35.94376325
川越市東部地域ふれあいセンター	埼玉県川越市並木452-1	139.5226993	35.9021551
川越駅西口第一自転車駐車場	埼玉県川越市脇田本町25-4	139.4810013	35.90891204
川越駅西口第二自転車駐車場	埼玉県川越市脇田本町39-66	139.4838061	35.9058895
川越駅東口自転車駐車場	埼玉県川越市菅原町22-12	139.4847141	35.90832521

図表6 AED 設置場所データ編集後の列

出典：【川越市】AED 設置場所（公共施設），<https://opendata.pref.saitama.lg.jp/data/dataset/aed>  
川越市保健医療部保健医療推進課作成 CCBY2.1jp <https://creativecommons.org/licenses/by/2.1/jp/>

### [手順3] データの読み込み，地図の描画

jSTAT MAP を起動し，川越市の AED 設置場所データと町丁・字等別人口密度を組み合わせた地図を作成する。ここでは，AED 設置場所データを先に描画する。

メニューの [統計地図作成] を選択するとサブメニューが表示される。その中から [プロット作成] を選択すると，プロットの作成方法を選択する画面が表示される。手順2で整形した CSV データを用いるので，[緯度経度付きファイル] を選択する。図表7のファイルを選択するウィンドウが表示されるので，[参照] ボタンを押し，手順2で整形したデータを読み込めた件数が表示されるので，必要に応じてグループ名やアイコンを変更して [登録する] ボタンをクリックする。以上で AED 設置場所が地図上に表示される。

緯度経度付きファイルから地図上にプロットを作成する方法です。  
読み込む緯度経度付きCSVファイル(.csv)を指定して下さい。

CSVファイル: kawagoe\_aed.csv 参照

文字コードはShift-JISです。UTF-8の場合は、UTF-8にチェックしてアップロードして下さい。  UTF-8

取り込むCSVファイルは以下のようなフォーマットとします。

店舗名	住所	電話番号	面積	駐車数	売上	URL	経度	緯度
S	S	S	N	N	N	U	N	N
新宿店	東京都新宿区〇〇	03-1234-5678	100	5	100	http://www...	139.691713	35.6894431
高田馬場店	東京都新宿区〇〇	03-1234-9876	120	10	96	http://www...	139.642324	35.4475001

属性項目の型を指定する場合は2行目に記載する。(省略可)

※世界測地系10進の経度緯度を記載下さい。  
※最初の属性が名称、末尾の2列が必ず、経度、緯度となります。  
※取り込み可能な件数は最大1000件であり、それを超えた分のデータは取り込まれません。  
※属性項目の型を指定する場合は、2行目を型指定の行とします。  
(型を指定しない場合は、経度、緯度以外は文字属性として取り込まれます)  
※文字属性は「S」、数値属性は「N」、URL属性は「U」を1行目の属性項目に対応させて指定して下さい。

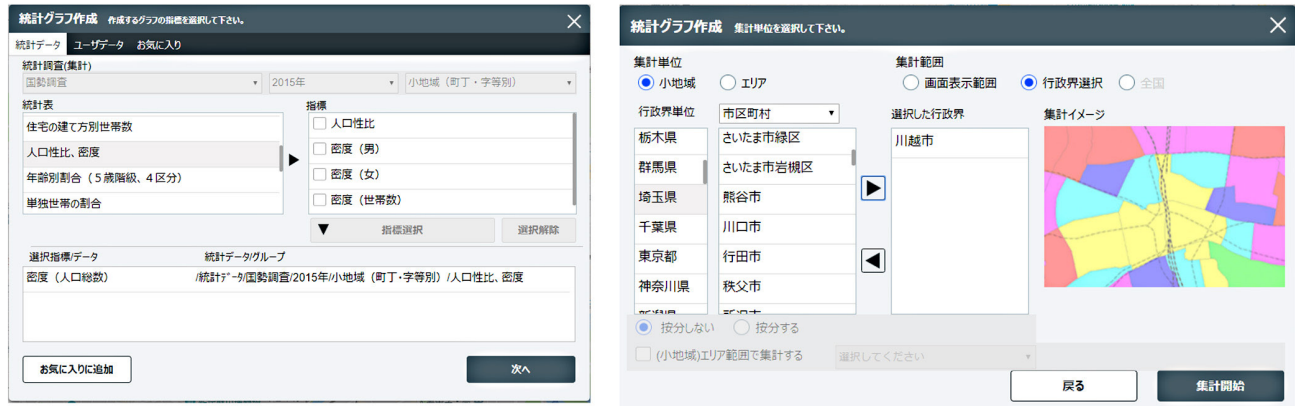
アップロードする

図表7 緯度経度付きの CSV データ選択画面

出典：政府統計の総合窓口 (e-Stat) (<https://www.e-stat.go.jp/>)

次に，川越市の小地域（町丁・字等別）の人口密度を描画する。人口密度のデータは，政府統計の総合窓口 (e-Stat) で公開されているデータを用いる。メニューの [統計地図作成] を選択し，サブメニューの中から [統計グラフ作成] を選択する。図表8の左図が表示されるので，調査名を [国勢調査]，年を [2015年]（最新のもの），集計単位を [小地域（町丁・字等別）]，統計表を [人口性比，密度]，指標を [密度（人口総数）] と順に選択する。表示する指標である [密度（人口総数）] を選択できたら，[指標選択] ボタンをクリックする。

図表8右図に切り替わるので，小計単位を [小地域]，集計範囲を [行政界選択] とすると，都道府県，市区町村の順に選択できる。ここでは，埼玉県川越市を選択する。目的の行政界を選択したら，[集計開始] ボタンをクリックする。



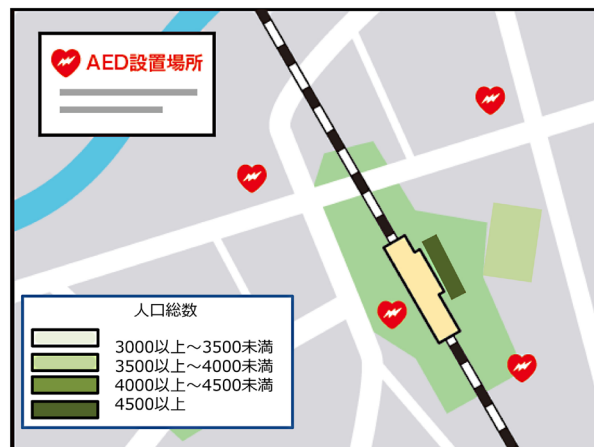
図表 8 人口密度を描画するためのメニュー画面

出典：政府統計の総合窓口 (e-Stat) (<https://www.e-stat.go.jp/>)

必要に応じて、人口密度の刻み幅を変更したり、凡例の表示を修正することにより、図表 9 のような地図が描画される。

#### [手順 4] 地図の分析

描画された図から、人口密度が高いにも関わらず AED の設置が少ない場所を見つける。



図表 9 人口密度と公共 AED 設置場所の地図の描画イメージ

#### <参考文献・参考サイト>

- ・「ICT 利活用の促進 オープンデータ戦略の推進」総務省,  
[http://www.soumu.go.jp/menu\\_seisaku/ictseisaku/ictriyu/opendata/](http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyu/opendata/)
- ・「オープンデータ化支援研修の教材～第 2 部オープンデータを公開するための手順」総務省,  
[https://www.opendata-training.org/wp-content/uploads/material/2\\_support/2\\_new/shien\\_opendata\\_tejun.pdf](https://www.opendata-training.org/wp-content/uploads/material/2_support/2_new/shien_opendata_tejun.pdf)
- ・「政府統計の総合窓口 (e-Stat)」<https://www.e-stat.go.jp/>
- ・「埼玉県オープンデータ」埼玉県, <https://opendata.pref.saitama.lg.jp/>
- ・「オープンデータ基本指針」総務省,  
<https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/kihonsisin.pdf>



## ■ 学習活動と展開

### 【学習活動の目的】

- 身近な情報システムにおけるデータの位置付けや、データを蓄積・管理・提供する仕組みと特徴について理解する。
- インターネット上で公開されているデータ、国や地方公共団体、民間企業が公開するオープンデータから必要なデータを集め、データを分析する力を身に付ける。
- 複数の情報システムが提供するサービスを比較検討し、目的に応じて最適なものを選択したり、組み合わせたりして活用する力を養う。
- データを多面的に精査して活用し、情報社会に主体的に参画しようとする態度を養う。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	コンビニエンスストアのレジで、どのようなデータを集めることができるか考えてみよう。	コンビニエンスストアで買い物をする際にどのようなデータが収集でき、どのように活用できるか考える。
展開 2	国や地方公共団体、民間企業が公開するオープンデータには、どのようなデータがあるか確かめてみよう。	国や地方公共団体、民間企業が公開するオープンデータを確かめ、収集する。
展開 3	オープンデータを組み合わせて、どのように問題発見や問題解決につながられるか考えてみよう。	収集したデータを組み合わせて表示し、どのような問題発見や問題解決に役立てることができるかを考える。

展開 1	
問 い	コンビニエンスストアのレジで、どのようなデータを集めることができるか考えてみよう。
学習活動	<ul style="list-style-type: none"> <li>• コンビニエンスストアでの買い物を基に、情報システムがどのようなデータを蓄積し提供されているか考える活動を通じて、データの蓄積や活用方法について考える。</li> <li>• 個人で考えた後、グループで議論を行い、考えを広げる。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>• 活動につまずいている生徒に必要なに応じてヒントを与え、生徒自らがデータの蓄積や活用方法に気づくよう促す。</li> </ul>



展開 2	
問 い	国や地方公共団体，民間企業が公開するオープンデータには，どのようなデータがあるか確かめてみよう。
学習活動	<ul style="list-style-type: none"> <li>国や地方公共団体，民間企業が公開するオープンデータから必要なデータを探し収集することを通して，どのようなデータが提供され，どのように活用できるか考える活動を行う。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>日常の中で考えている問題を引き出したり，提供されているデータの関連性についてヒントを与えたりして，生徒自らが必要なデータについて気づくよう促す。</li> </ul>



展開 3	
問 い	オープンデータを組み合わせると，どのように問題発見や問題解決につながられるか考えてみよう。
学習活動	<ul style="list-style-type: none"> <li>複数のデータを組み合わせるとして分析する活動を通じて，問題の発見や問題解決に，発信者の意図を読み取る力を身に付ける。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>データを組み合わせるとして分析した際に，問題発見や問題解決につながらない場合でも，データを組み合わせるとして分析したことにより発見できたことがあることに気づかせる。</li> </ul>



まとめ	
まとめ	<ul style="list-style-type: none"> <li>データを効果的に活用可能にするためには，情報システムによる望ましいデータの蓄積・管理・提供が必要になることをまとめる。</li> </ul>

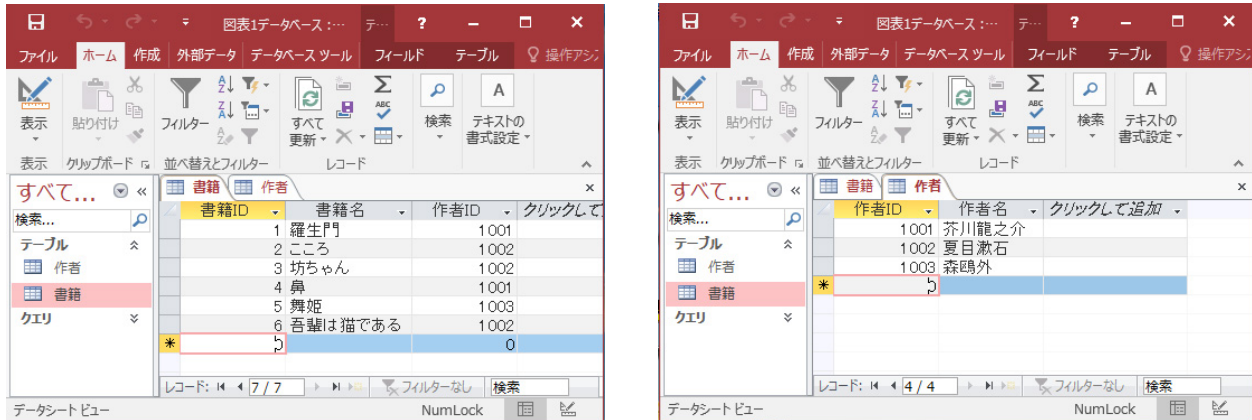
## ■研修内容

### 【研修の目的】

- リレーショナルデータベースでのデータの操作を理解する
- つながりを表すなど、表形式以外でのデータの表現形式を理解する
- 項目（キー）と値（バリュー）をセットにしたキー・バリュー形式のデータを扱うための方法について理解する
- API を用いてデータを取得する方法について理解する
- キー・バリュー形式で表現されたデータを活用して、社会的な問題を発見する授業ができるようになる

### (1) リレーショナルデータベース

データの蓄積や検索が容易にできるよう、一定の形式でデータを整理してデータベースが用いられている。データベースのうち、特にリレーショナルデータベース（関係データベース）が広く用いられている。リレーショナルデータベースは、表形式で表現されたテーブルというデータの集合として管理する。テーブルの列には、フィールドといわれる「書籍名」や「作者名」という項目を割り当てる。1件分のデータはレコードといわれる1つの行で表す。また、複数のテーブルを関連付けて、結合して利用することもできる。リレーショナルデータベースには、有料のもの他に MySQL, PostgreSQL などの無料で使えるものがある。リレーショナルデータベースを操作するには、SQL(Structured Query Language) という言語が用いられる。リレーショナルデータベースについての例を図表 1 に示す。



図表 1 リレーショナルデータベースの例

リレーショナルデータベースでは、「選択」・「射影」・「結合」という関係演算を用いて処理することが多い。選択はテーブルから条件を満たすレコードを抽出する操作である。図表 1 では、書籍テーブルから作者 ID が 1001 であるレコードを選択した場合を示している。射影はテーブルから一部のフィールドを取り出す操作である。図表 1 では、書籍テーブルから「書籍名」のフィールドを抽出する例を示している。結合は複数のテーブルを 1 つの表にする操作である。

テーブルに対して、データを追加したり、レコードを削除したりすることもできる。これらの操作をする SQL も演習 1 に示す。

例に示したデータベースは、書籍と作者の 2 つのテーブルで構成されている。1 つのテーブルにまとめてデータを管理することもできるが、すべてのレコードに同じ作者名が含まれることになる。テーブルを分けることにより、データが冗長になってしまうことを避けることができる。また、作者名の値を変更しなければならなくなった場合に作者名テーブルを更新するだけでよく、書籍テーブルを変更することなく済ませることができ変更漏れも生じない。このようなテーブル設計の工夫を、テーブルの正規化という。



## <演習 1>

リレーショナルデータベースを使って、簡単な蔵書目録を作りましょう。また、SQL を使ってデータベースから次の操作を試みましょう。

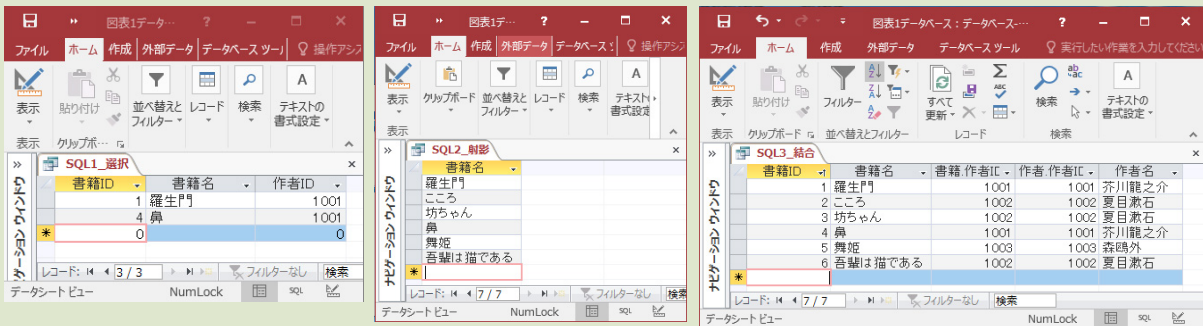
- ① 特定の作者の書籍データを表示する
- ② 所蔵している蔵書の作者名の一覧を表示する
- ③ 所蔵している蔵書の書籍と作者が一つの表になった形で表示する

SQL の例

書籍のテーブルから作者 ID が 1001 であるレコードを選択する  
`SELECT * FROM 書籍 WHERE 作者 ID=1001;`

書籍テーブルから書籍名のフィールドを抽出する (射影)  
`SELECT 書籍名 FROM 書籍;`

書籍テーブルと作者テーブルで作者 ID が等しいレコードを結合する  
`SELECT * FROM 書籍 INNER JOIN 作者  
ON 書籍 . 作者 ID = 作者 . 作者 ID;`



書籍テーブルに新しいレコードを追加する

```
INSERT INTO 書籍 (書籍 ID, 書籍名, 作者 ID)
VALUES (7, '高瀬舟', 1003);
```

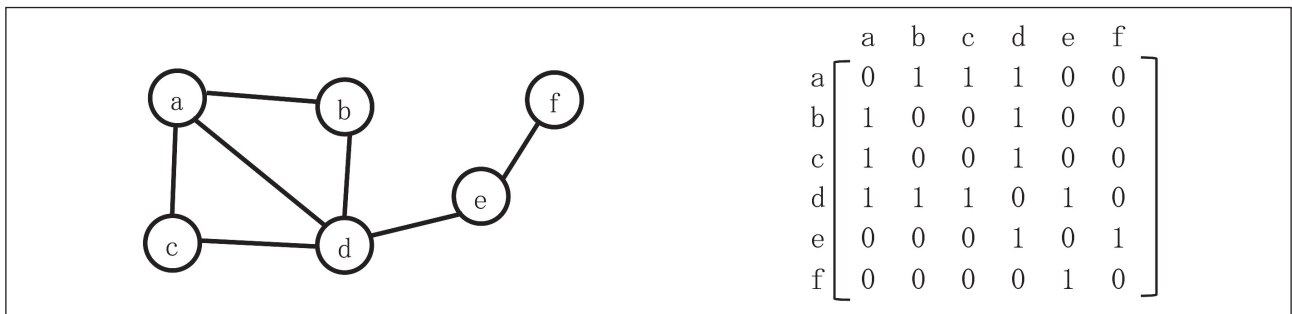
書籍テーブルからレコードを削除する

```
DELETE FROM 書籍 WHERE 書籍 ID = 7;
```

## (2) データのさまざまな表現形式

つながりを表すデータなど、表形式以外の方法でデータを表現した方が扱いやすい場合がある。そのようなデータを表現する方法を理解する必要がある。

SNS での人のつながりや鉄道の路線などのような関係性について図を使って表現する。図表 2 では、a から f までの 6 人の人のつながりを表している。a から f までの人を表す丸を頂点 (バーテックスまたはノード)、人と人とのつながりを表す線分を辺 (エッジ) といい、頂点と辺で構成された図を離散グラフという。図表 2 では特に方向性を考えない無向グラフになっている。人のつながりでの片想いの関係や一方通行の道路のような、方向性を考慮した有向グラフを考えることもできる。



図表 2 人のつながりを表現する離散グラフ (左) と隣接行列 (右)

図表 2 の離散グラフをコンピュータ上で表現するには、コンピュータで扱うことができるデータ形式で表現する必要がある。一つの例として、2次元配列を用いて表現することができる。頂点どうしがつながっていれば 1,

つながっていないければ 0 として表現する。このように表現する方法を隣接行列という。隣接行列を用いて表現する場合、辺の本数が少ない場合には、隣接行列の要素の多くが 0 である疎行列になってしまう。

そこで、接続している頂点のリストを頂点ごとにデータとして、次のような隣接リストとして表現する方法が考えられる。頂点 a には頂点 b と c が接続していることを示し、頂点 b 以下も同様である。

```
{a:[b,c,d], b:[a,d], c:[a,d], d:[a,b,c,e], e:[d,f], f:[e]}
```

隣接リストでは、注目している頂点という項目と、接続している頂点のリストという値との組として表現されている。このように、項目(キー)と値(バリュー)との組により表現されるデータの形式をキー・バリュー形式という。

隣接リストの形式では、疎行列のような不要なデータについては解消されるが、頂点同士が接続しているかどうかを調べるには隣接リストをたどって調べる必要が生じる。離散グラフをコンピュータ上で扱う際には、離散グラフを用いてどのような問題を解決するかを考えて表現することが求められる。

### (3) Web API によるデータの取得

インターネット上のシステムによりさまざまなデータが提供されている。前項ではサイト上から csv 形式のデータをダウンロードして問題解決に用いたが、プログラムでデータを扱うよう Web API (Application Programming Interface) によりデータを提供しているサイトが多くある。Web API では、XML 形式や JSON 形式でデータが提供されている場合が多い。XML 形式での属性と要素の組合せ、JSON 形式でのキーと値の組合せにより、項目と値をセットにして扱うことができる。

ここで、実際に Web API を利用してデータを取得する。Web API を利用するには、一般にユーザ登録が必要である。

ここでは例として、YAHOO!JAPAN デベロッパーネットワークに登録し、気象情報 API を使って、JSON 形式と XML 形式のデータを取得する。

YAHOO!JAPAN のユーザ登録をし、ログインしてクライアント ID の申請を行う。API を使えるようにするため、「新しいアプリケーションを開発」の登録を <https://e.developer.yahoo.co.jp/register> にアクセスして行う。クライアント ID を取得できたら、ブラウザから Web API を利用して 60 分後までの降水量の予測値のデータを XML 形式と JSON 形式で取得する。ここでは、取得できるデータ形式の一部を図表 3 に示す。「クライアント ID」と書かれた箇所には、先に登録したクライアント ID を用いる。

データを取得する URL (入力時には改行しない)

```
https://map.yahooapis.jp/weather/V1/place?appid={クライアント ID}
&coordinates=経度(東経),緯度(北緯)&output={json または xml}
```

【XML 形式】	【JSON 形式のデータ】
<pre>&lt;Weather&gt;   &lt;Type&gt;observation&lt;/Type&gt;   &lt;Date&gt;201901061950&lt;/Date&gt;   &lt;Rainfall&gt;0.00&lt;/Rainfall&gt; &lt;/Weather&gt;</pre>	<pre>"Weather": {   "Type": "observation",   "Date": "201901061950",   "Rainfall": 0.00 }</pre>

図表 3 取得したデータ (XML 形式および JSON 形式の一部)

XML 形式では、タグがキーになっており、タグ間にバリューがはさまれている。JSON 形式では、コロンの前が項目、コロンの後ろが値になっている。若干の値の違いはあるが、いずれも項目と値をセットにしてデータが表現されている。API については、サービスが停止したり、該当の URL が変更になったり、問い合わせの書式が変わったりする場合がある。API を使ってデータを取得する場合は、そのサービスが提供されているかを確認するとともに、最新の URL と問い合わせの書式を調べて行う必要がある。

### (4) キー・バリュー形式のデータの処理・蓄積

キー・バリュー形式のデータを扱うことができるプログラミング言語もある。プログラミング言語によっては、連想配列・ハッシュ・辞書などと呼ばれている。ここでは、Python の辞書というデータ構造を用いて連絡先のリストを作ることを題材として、データを追加し、表示するプログラムを扱う。

携帯電話の普及に伴い、固定電話を自宅に有しない人もいる。また、一人で複数のメールアドレスを使っている人も珍しくない。次にあげる 3 名の連絡先をキー・バリュー形式で members という配列に代入し、表示するプログラムを例示する。

▲登録するデータ

A 田さん 自宅電話 homephone:0\*\*-\*\*\*-1111  
携帯メール mobilemail:ata@\*\*\*.ne.jp  
B 川さん 携帯電話 mobilephone:080-\*\*\*\*-2222  
C 山さん 携帯電話 mobilephone:090-\*\*\*\*-3333  
PCメール email:cyama@\*\*\*.com

```
1 # 初めから members に登録する場合
2 members = [ { "name" : "A 田" ,
3               "homephone" : "0**-***-1111" ,
4               "mobilemail" : "ata@***.ne.jp"
5             } ,
6             { "name" : "B 川" ,
7               "mobilephone" : "0**-***-2222"
8             }
9           ]
10
11 # 後から members に追加する場合
12 members.append( { "name" : "C 山" ,
13                  "mobilephone" : "090-****-3333" ,
14                  "email" : "cyama@***.com"
15                } )
16
17 # すべてのデータを表示する
18 for member in members:
19     for key , value in member.items():
20         print( key + ' : ' + value )
21     print()
```

図表 4 キーとバリューで扱うプログラム (Python)

(1) でリレーショナルデータベースを扱ったが、Amazon Dynamo DB やオープンソースである Hibari といったキー・バリュー形式に対応したデータベースもある。また、他のデータ形式に対応したデータベースもあり、これらのデータベースを総称して NoSQL と呼ばれている。NoSQL はビッグデータの活用と関連して注目されている。

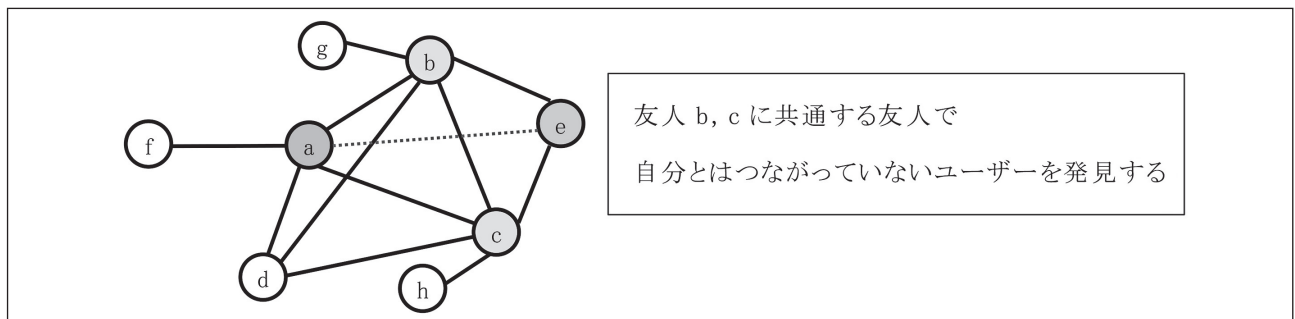
<演習 2>

1~2名程度のデータを、上の Python のプログラムに追加してみましょう。

(5) キー・バリュー形式のデータを用いた問題発見

ここで、キー・バリュー形式で取得できたデータを用いて、問題発見に役立てる。SNS の開発者登録が必要ではあるが、SNS でつながりがあるユーザーのデータを Web API を利用して取得することができる。ここでは、Web API を利用して、JSON 形式のデータを取得できたものとして演習を行う。

下の図表 5 で自分を a とする。自分の 2 人の友人 b, c が共通につながっている人は、自分の友人である可能性が高い。このような人を発見して、おすすめのユーザーとして提案する。



図表 5 SNS でのつながりのデータからおすすめのユーザーを見つける

上の図表 5 で、自分 a のつながりがあるユーザーは、次のような隣接リストで表現できる。

```
自分 {"ids":["b","c","d","f"]}
友人 b, c についても同様に隣接リストはそれぞれ
友人 b {"ids":["a","c","d","e","g"]}
友人 c {"ids":["a","b","d","e","h"]}
```

とである。SNS での隣接リストは、Web API を利用すると同様のデータが JSON 形式で取得できる。取得したデータから、次のように提案する。

- 1) 友人 b, c の共通の友人を見つける。共通の友人は, ["a","d","e"] である。
- 2) ここから、自分自身である a と、既に友人となっている d を除く。
- 3) その結果である ["e"] を共通の友人として提案する。

プログラムを使って、同様の処理を行うことができる。

### <演習 3>

解答例のフォルダの内容を表示し、プログラムを実施して動作を確認するとともに、プログラムリストを見て動作を確認してみましょう。

### <参考文献・参考サイト>

- ・「Yahoo!JAPAN デベロッパーネットワーク」ヤフー（株），<https://developer.yahoo.co.jp/>



## ■ 学習活動と展開

### 【学習活動の目的】

- リレーショナルデータベースでのデータの操作を理解する。
- つながりを表すなど、表形式以外でのデータの表現形式を理解する。
- 項目（キー）と値（バリュー）をセットにしたキー・バリュー形式のデータを扱うための方法を理解する、
- APIを用いてデータを取得する方法について理解する。
- データの特性に応じて、データの形式の表現方法について思考して選択したり、さまざまな形式によりデータ表現したりする力を養う。
- さまざまな形式のデータを活用し、情報社会に主体的に参画しようとする態度を養う。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	リレーショナルデータベースを用いて、蔵書管理をしよう。	書籍データについて、書籍の検索をしたり、書籍データを追加したりして、リレーショナルデータベースを操作する。
展開 2	人と人とのつながりを表現する方法を考えてみよう。	人と人とのつながりについて、離散グラフや隣接行列、隣接リストといった方法で表現する。
展開 3	キー・バリュー形式で表現されたデータからおすすめのユーザーを探そう。	キー・バリュー形式で表現されたデータを読み取ったり、プログラムで扱ったりする。

#### 展開 1

問 い	リレーショナルデータベースを用いて蔵書管理をしよう。
学習活動	・ 蔵書リストを題材にして、矩形形式で扱われるデータ表現とその操作方法について考える。
指導上の留意点	・ 活動につまずいている生徒に、必要に応じてヒントを与え、生徒が操作を理解しながらデータベースを活用できるよう促す。



展開 2	
問 い	人と人のつながりを表す方法を考えてみよう。
学習活動	<ul style="list-style-type: none"> <li>・ 離散グラフにより表現されたつながりに対し、隣接行列、隣接リストにより表現する。</li> <li>・ 隣接行列、隣接リストでのつながりの表現を見て、それぞれの特長について個人で考える。その後、グループで議論を行い、考えを広げる。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>・ つながりがある関係性を実際に表現したり、つながりをたどったりして、データの表現のしやすさや扱いやすさを比較するように生徒に促す。</li> </ul>



展開 3	
問 い	キー・バリュー形式で表現されたデータからおすすめのユーザーを探そう。
学習活動	<ul style="list-style-type: none"> <li>・ キー・バリュー形式で表現されたデータを図示したり、プログラムで扱ったりする活動を通して問題解決に用いる。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>・ キー・バリュー形式のデータと、それが表すものとの関係が理解できるようにする。</li> <li>・ キー・バリュー形式で表現されたデータを、プログラムで扱う方法を理解できるようにし、そのデータを用いて問題解決に有用であることが理解できるようにする。</li> </ul>



まとめ	
まとめ	<ul style="list-style-type: none"> <li>・ データの特徴に応じてデータを表現する必要があることを理解し、それぞれのデータの表現形式について整理する。</li> </ul>

## ■研修内容

### 【研修の目的】

- 名義尺度，順序尺度，間隔尺度，比例尺度などのデータの尺度水準の違いを理解する。
- 量的データについて，データ処理の流れ及びその評価について理解する。
- データを問題の発見・解決に活用するために，必要なデータの収集について，選択，判断する力を養う。
- データを多面的に精査して活用し，情報社会に主体的に参画する態度を養う授業ができるようになる。

### (1) 量的データと質的データ

例えば，自分が住んでいる地域の夏の気象の傾向について調べるため，気象庁のウェブサイトから図表1のような日別のデータ(図表1-A)を，また，地域で取引される中古住宅の価格帯の傾向を調べるため，不動産会社のウェブサイトからデータ(図表1-B)を取得したとする。

月日	最高気温	最大風速	風向
7月1日	24.3°C	6.1m/s	東南東
7月2日	28.1°C	5.2m/s	東北東
・	・	・	・

図表 1-A 日別の気象データ

住宅番号	価格 (百万円)	土地 (m <sup>2</sup> )	建物面積 (m <sup>2</sup> )	築年数	車庫	住宅の 構造	内装状況
1	25.8	86.4	66.0	48	0	木造	やや良好
2	25.8	115.0	83.4	17	0	木造	良好
3	27.8	73.1	58.3	4	1	鉄筋	極上
4	29.8	103.5	89.3	32	0	木造	良好
・	・	・	・	・	・	・	・

図表 1-B 中古住宅の取引価格等データ

データから対象の傾向を分析する場合，1行目にデータの項目名，2行目以降にその項目の値が入力された表を作成する(データの整理)。このように行が観測対象，列が観測項目に対応したデータで表現されるデータセットを構造化データという。

上記の気象データ(図表1-A)は時点が対象となり，その系列に沿って得られる時系列データといい，中古住宅の取引価格等データ(図表1-B)は時点が固定されて同質の対象(中古住宅)が複数観測されるクロスセクションデータという。どちらも，対象を大量に観察することで対象の傾向を分析することができる。時系列データは時間軸に沿った傾向(トレンド)の分析も行うことができる。

ここで記録されているデータの項目は，風向き，住宅の構造，内装状況のように分類項目(カテゴリー)の対象によってばらつく質的データと数直線上の値(数量)としてばらつく量的データに大きく分類される。さらにデータは測定される尺度(モノサシ)の違いで次の4種類に分けることができる。

質的データとして①名義尺度，②順序尺度，量的データとして③間隔尺度，④比例尺度などがある。①名義尺度の例としては，性別などがあり，対象について男，女などの分類をすることができる。②順序尺度の例としては，成績評価などがあり，対象について優，良，可のように分類をできるとともに，優>良>可のような順序構造がある。ただし，優と良，良と可の程度の差が一致する保障はない。③間隔尺度の例としては気温などがあげられる。摂氏と華氏の0が違うなど絶対的な原点はないが等単位制があるので，データの差は意味を持ち，データの比は意味を持たない。④比例尺度の例としては，身長や体重などがある。0cmや0kgなど，絶対的な0を持ち等単位制もあるので，データの差だけでなくデータの比も意味を持つ。

尺度が異なれば，分析処理やグラフ表現に違いが出てくる。下記は尺度と分析の方法があっていない例である。

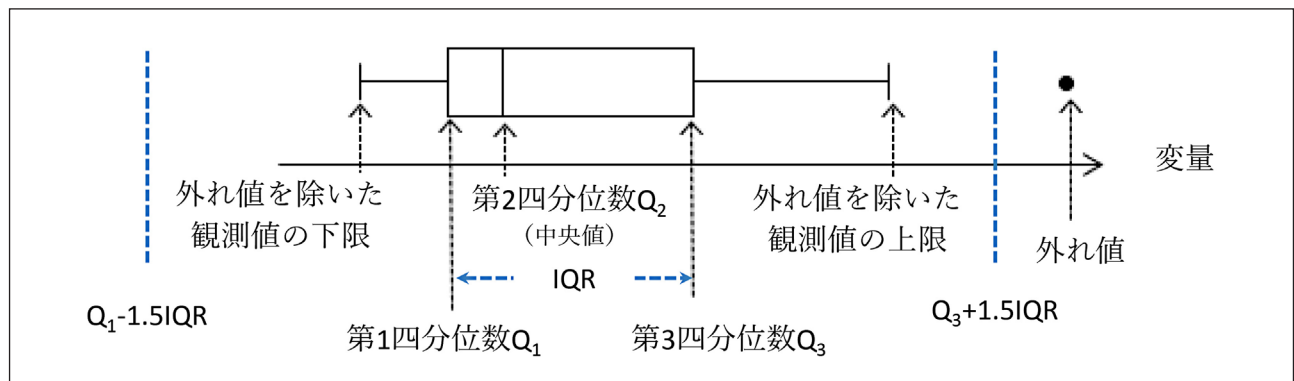
- ・ 数字で入力されている名義尺度のデータについて平均を計算してしまう。
- ・ 順序尺度のデータをカテゴリー別に集計した結果をカテゴリーの順序を無視して棒グラフや円グラフに表してしまう。(カテゴリーの例として，優，良，可など)
- ・ 間隔尺度のデータに対して，例えば，気温15°Cの地点Aと気温30°Cの地点Bについて，地点Bは地点Aより2倍暑いと判断してしまう。



データ分析では、記録されたデータをそのまま統計処理するだけではなく、適時、欠測値の処理や層別、変数変換、水準化などのデータ変容（マネジメント）を行う。対象によっては、すべてのデータ項目が記録されていない場合もある。そのような記録のないデータを欠測値という。欠測値は、何も記入しないでおく、もしくは欠測コード（例えば 999）を入力してデータとして区別できるようにしておく。その際、空白が数値の 0 や欠測コードが数値と混同して演算されないように注意する。欠測値は分析に際して、（ア）欠測値を含む対象全体の行（レコード）を削除する、（イ）欠測値を近いと思われる値で埋める等の処理を行うが、分析結果を提示する際にはその処理を明記する必要がある。

層別とは、名義尺度や順序尺度で記録されたデータの値を基に、データの対象をグループに分けて比較分析等を行うことを指す。風向別に気温の傾向を比較、住宅構造別に価格の傾向を比較する等である。変数変換とはデータ項目の変換処理のことを言い、例えば、県別データ等で人口規模の影響を取り除いて比較するために人口 1000 人あたりの県民所得や病院数を算出したり、各住宅に対して床面積 1 m<sup>2</sup>あたりの価格に変換したりするなどのデータ項目の変換処理をいう。水準化（レベル化）とは、量的データがある値でいくつかの区切って、順序尺度（質的データ）に変換することをいう。気温で、対象とする日を普通、夏日、真夏日などとする、世帯別所得金額を基に高所得、中所得、低所得と階級にする等である。

ここでは量的データの傾向を分析する方法を紹介する。学習指導要領では、小学校で、量的データのばらつき（分布）の記述として、ドットプロットや中心位置を示す統計量（代表値）である平均値・中央値・最頻値を学習する。ただし、最頻値に関しては、離散型データの場合は最頻値、連続型データの場合は階級を取った際の最大度数を示す最頻階級もしくはその階級値を指す。中学校では、度数分布やヒストグラム、四分位数及び四分位数に最小値と最大値を加えた 5 つの位置統計量（5 数要約）を示す箱ひげ図による分布の意味を学習する。高等学校の数学 I では、分布の広がりを示す統計量である分散・標準偏差、外れ値や統計的仮説検定の考え方について学習する。四分位数はデータを大きさの順に並べたときに、データの個数（相対度数）を 4 等分する境界値で、データが小さい方から順に第 1 四分位数（Q1）、第 2 四分位数（Q2）（中央値と同値）、第 3 四分位数（Q3）という。累積相対度数は中学校で学習するが、四分位数はそれぞれ、下側累積 25% 点、下側累積 50% 点、下側累積 75% 点に相当する。四分位数を図示する方法として、箱ひげ図が用いられる。中学では四分位数に最小値と最大値を加えた 5 つの位置統計量（5 数要約）を示す箱ひげ図、高校数学 I では、第 1 四分位数から第 1 四分位数 - 1.5 × 四分位範囲（IQR）までの間の最小値、第 3 四分位数から第 3 四分位数 + 1.5 × 四分位範囲（箱の長さ）までの間の最大値をひげの両端とし、それ以上離れた値を外れ値として点で示す箱ひげ図（図表 2）を学習する。



図表 2 箱ひげ図（外れ値を表示）

箱ひげ図は、箱の部分が分布の中心傾向を示し、ひげの部分が分布の裾を表す。箱ひげ図を並べることで量的データの分布の比較が簡単にできる（並列箱ひげ図）。

### <演習 1>

2018 年 7 月 23 日に、熊谷は日本で最高気温となる 41.1℃を記録しました。この年の夏のニュースでは、連日熱中症で搬送される人数が伝えられました。しかし、2018 年の熊谷は本当に暑かったのでしょうか？ 2017 年と 2018 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。

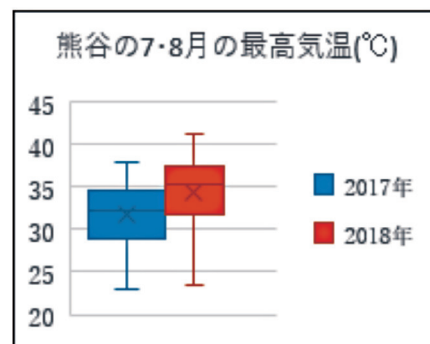
#### ■具体的な手順

- 1) 気象庁のサイトから、2017 年および 2018 年の 7 月 1 日から 8 月 31 日の熊谷の日最高気温のデータをダウンロードする。
- 2) 表計算ソフトウェアでダウンロードしたデータを開く。表計算ソフトウェアを用いて、最小・第 1 四分位数・中央値・第 3 四分位数・最大値・平均値・分散・標準偏差を求める。これらを求める関数は、図表 3 である。

これらを求める関数を、図表3に示す。

- 3) 求めたデータをもとに分析する。2年間の統計量を比較すると、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2018年の方が高くなっている。このことから、2017年と2018年の熊谷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。

統計量	2017年	2018年	用いる関数
最高	37.8	41.1	=MAX(範囲)
第3四分位数	34.6	37.4	=QUARTILE.INC(範囲,3)
中央値	32.2	35.4	=MEDIAN(範囲)
第1四分位数	29.1	31.9	=QUARTILE.INC(範囲,1)
最低	23.0	23.4	=MIN(範囲)
平均値	31.8	35.9	=AVERAGE(範囲)
分散	13.1	15.9	=VAR.P(範囲)
標準偏差	3.6	4.0	=STDEV.P(範囲)



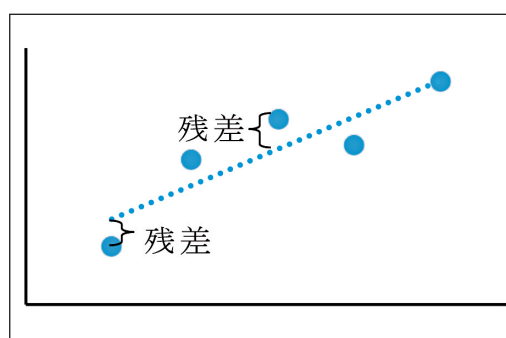
図表3 統計量を求めた結果と用いる関数・箱ひげ図(5数要約)

## (2) 量的データ間の関係

次に、(x,y)の関係性を数学Iで学習する散布図・相関係数によって分析する。散布図上のデータのパターンから、右肩上がりの直線傾向(変数xの値が増えたとき変数yの値が直線的に増える傾向)にあるとき、2つの量的データには正の相関関係があるという。逆に変数xの値が増えたとき変数yの値が直線的に減る傾向にあるとき、負の相関関係があるという。相関関係があれば、次節で述べる単回帰直線(変数xの値が与えられたときの変数yの条件付平均値を表す直線)で、yの予測モデルを作成することができる。例えば、図表1-B 中古住宅の取引価格等のデータで、土地や床面積の広さから価格を予測する直線のモデル式を推測することができる。2変量間の関係は相関関係だけではなく、曲線的なパターンが当てはまることもある。散布図のデータに特段のパターンがない場合、言い換えれば変数xの値が変化しても変数yの分布に違いが見られない場合は、変数xとyの間に関係がない。つまり変数xは変数yの変化を予測するモデル構築に役に立たないということになる。相関関係の強弱と向きは、相関係数で評価できる。相関係数は-1から1の値を取り、符号が直線傾向の傾きに対応する。また、データが直線に近づいて分布するほど、絶対値の大きさが1に近くなる。表計算ソフトウェアのExcelではCORREL関数によって相関係数を求めることができる。

## (3) 単回帰分析を用いた値の推測

中古住宅の例では、住宅の価格の違いを住宅が持つ土地や床面積などの属性の違いで説明する予測式を導くことが分析の目的となる。このとき、予測される変数を目的変数といい、散布図ではy軸になる。また予測に用いる変数を説明変数といい、x軸に対応させる。データ(x,y)の間に相関関係があれば、直線(一次関数) $y = \alpha x + \beta$ で傾向線を当てはめることができる。この直線を単回帰直線という。データが与えられたとき、データに最も当てはまるように直線の傾き(回帰係数)



残差を2乗した値の総和が最小になるような回帰直線の決定方法を最小二乗法という。

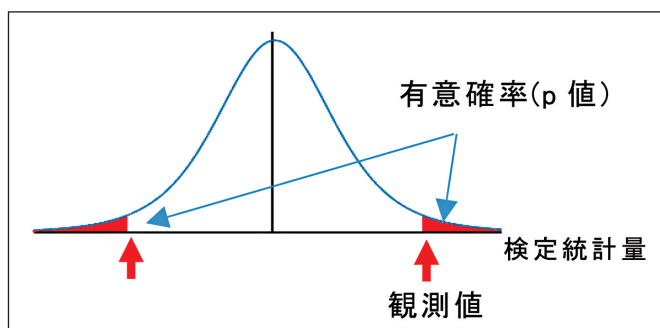
図表4 回帰直線と残差

や切片を求め、傾きの解釈やyの予測を行うことを単回帰分析という。xが与えられたときの直線上の値(モデル値)と実際のyの値との差を残差といい、回帰直線は残差を2乗した値の総和が最小になるように決定されている(最小二乗法)。残差の絶対値が小さいほど回帰直線(モデル)の説明力が高いことになる。このモデルの説明力を測る指標が決定係数 $R^2$ で、0から1の範囲の値をとる。決定係数の値が1に近いほど、回帰直線のデータへのあてはまりがよいといえる。データ点がすべて直線上にのっている状態、残差がすべて0となる状態では決定係数1となる。

## (4) 量的データの統計的仮説検定

量的データは、個数などの0, 1, 2, ...という整数で表される離散型データと、長さ・重さなどの実数値で表される連続型データに分類される。データが研究対象である母集団から抜き取られた一部分(標本)から得られた調査データや実験データの場合、わたしたちが知りたいのはデータの背景である母集団特性値(母数)である。

母数として典型的なものに、離散型データに対する母集団での比率(母比率)である支持率など、連続型データに対する母集団での平均値(母平均)である平均所得などがある。統計的推測とは、未知の母数を全数調査することなく、標本データから推測する統計的な考え方のことであり、母数を点推定や区間推定する推定の方法と母数に関する仮説の真偽を検証する仮説検定(以下、「検定」という)に分けられる。



検定の考え方は、自動レコメンドや、Webページなどである一定期間にAとBの2パターンを用意しておき、どちらが効果的な結果を残したかによってWebページを自動改善するA/Bテスト等、近年のデータに基づく情報技術を理解する上で重要になっている。中学校では標本調査、高校数学Iでは仮説検定の考え方を学習する。検定は次の手順で行う。検証したい仮説(効果がある、差がある、等)を対立仮説 $H_1$ と呼び、その反対の仮説(効果がない、差がない、等)を帰無仮説 $H_0$ と呼ぶ。そして、標本データが帰無仮説上でどのような起こりやすさ(分布)の下で起こっているのかを確率で評価し、非常に小さな確率でしか起こらないことを証拠に、帰無仮説を棄却し対立仮説の成立を主張する。標本データから計算される検定統計量とその分布の違いによって、正規検定(Z検定)、t検定、 $\chi^2$ 検定等の検定の種類がある。また、対立仮説の方向によって、両側検定、右(左)片側検定を選択する。検定には帰無仮説が正しいのに棄却してしまう誤り(第1種の過誤 $\alpha$ )、対立仮説が正しく帰無仮説が誤りなのに棄却されない誤り(第2種の過誤 $\beta$ )の2種類の誤りがある。第1種の過誤 $\alpha$ を予め有意水準(検定の危険率)として、5%または1%で定めておき、帰無仮説上の分布で検定統計量の観測値が対立仮説の方向にその値以上にずれる確率(有意確率、p値)が有意水準以下になると、「帰無仮説は正しくない」と考え、帰無仮説は棄却され対立仮説を採択する。p値が有意水準より大きくなる場合は、帰無仮説が正しいというより、検出力の問題もあるので、今回のデータは帰無仮説を棄却する統計的な証拠にはならないという判断に留めることが重要である。

検定の目的とデータの前提条件等に応じて適切に検定の種類を選択すれば、データからエクセル関数等を利用して検定の有意確率を求めることができるので、検定の実行は容易である。

図表5 帰無仮説上での検定統計の分布と有意確率(両側検定)

## <演習2>

文部科学省・スポーツ庁では、平成10年度から継続的に握力・上体起こし・長座体前屈・反復横とび・20mシャトルラン・持久走・50m走・立ち幅とび・ハンドボール投げ・新体力テストの合計点の統計を取っています。これについて、体力・運動能力について関係が深い項目を見つけてみましょう。具体的には、16歳(高校1年生)の男子を対象に、これらの項目間の関係がどのようになっているかを図表6の手順で調べてみましょう。(政府統計の総合窓口(e-Stat)からデータを取得)。

次に、国立教育政策研究所が提供する<理科ねっとわーく>、小・中・高等学校向けの理科教育・学習用デジタル教材『科学の道工具箱』のサイトから同様に体力測定データを取得し、16歳(高校1年生)の男子を対象に、これらの項目間の関係がどのようになっているかを調べてみて結果の違いを考察してみましょう。

データの収集	<ul style="list-style-type: none"> <li>政府統計の総合窓口(e-Stat)からデータを検索・取得</li> <li>科学の道工具箱 <a href="https://rika-net.com/contents/cp0530/contents/08.html">https://rika-net.com/contents/cp0530/contents/08.html</a></li> </ul>
データの整形	<ul style="list-style-type: none"> <li>必要なデータを抽出する</li> <li>データを整形する</li> </ul>
データの分析	<ul style="list-style-type: none"> <li>散布図行列を描画して散布図や相関係数から関係を読み取る</li> <li>関係が深い項目を用いて、単回帰分析を行う</li> </ul>

図表6 演習の手順

### ■手順1 データ収集

政府統計の総合窓口(e-Stat)の、次のURLにある平成29年度体力・運動能力調査のテスト項目の年次推移のデータをダウンロードする。

<https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00402102&tstat=000001088875&cycle=0&tclass1=000001119456&second=1&second2=1>

※「体力運動能力調査 政府統計」で検索し、表示されたページで「平成29年度」を選択し、さらに「テスト項目の年次推移」を選択しても良い。

「科学の道工具箱」の中から「データライブラリー」のページから、「体力測定データ」を選択、その中の「02.高等学校体力測定データ」をダウンロードする。

<https://rika-net.com/contents/cp0530/contents/04-03-01.html>

## ■手順2 データ整形

### 【政府統計の総合窓口 (e-Stat)】

ここでは、16歳男子を対象に分析したいので、それぞれのテスト項目から16歳男子のデータを抽出する。調査項目によっては平成10年度よりも前からのデータも含まれているため、平成10年度以降のものだけに揃える。次の手順で散布図行列を描くため、各項目が列になるように、貼り付ける操作をする際には、「形式を選択して貼り付け」のうち「行/列の入れ替え」を用いて行う。

### 【科学の工具箱】

高等学校体力測定データには、高1から高3までの男女のデータが含まれている。その中から、「フィルター」もしくは「並べ替え」機能を使って、高1男子のみのデータとする。

## ■手順3-1 散布図行列を用いて散布図や相関係数から関係を読み取る

表計算ソフトウェアのExcelでは、2つの項目間の関係を散布図で表したり、相関係数を求めたりすることができる。しかし、一度に複数の項目間の散布図を描くことはできない。複数の2項目間の散布図を行と列に対応させて並べて描画したものを散布図行列、同様に相関係数を行列で表したものを相関係数行列という。散布図行列を描画できるExcelのマクロを用意した。それにより描画した散布図行列は図表8である。相関係数行列はエクセルの「データ」→「データの分析」→「相関」からも作成することができる。また、「科学の工具箱」の「統計グラフソフトを使う」からダウンロードできる中学生・高校生向けグラフソフトでも散布図行列の作成が可能である。

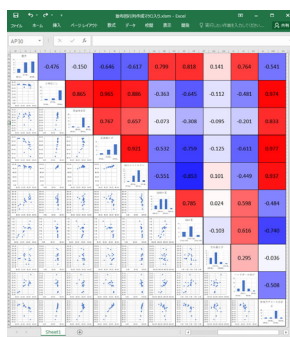
### 【政府統計の総合窓口 (e-Stat) からの時系列データ】

年度	握力	上体起こし	長座体前屈	反復横跳び	20mシャトルラン	1500m走	50m走	立ち幅跳び	ハンドボール投げ	新体力テストの合計
10	42.38	25.46	44.43	48.68	77.56	378.35	7.51	223.48	26.35	48.00
11	41.26	26.84	46.25	50.61	82.71	376.45	7.42	224.32	26.65	49.80
12	42.43	28.37	47.84	51.94	86.23	369.70	7.38	226.05	27.08	51.83
13	42.12	28.79	48.68	52.63	86.25	381.26	7.40	224.69	26.35	52.23
14	42.33	29.69	49.67	53.56	87.87	376.25	7.41	224.42	26.42	53.07
15	42.18	29.60	50.17	54.13	85.63	375.94	7.40	223.90	26.40	53.43

### 【科学の工具箱からの個人の体力測定データ】

性	身長	体重	座高	握力	上体起こし	長座体前屈	反復横跳び	シャトルラン	50m走	立ち幅跳び	ハンドボール投げ
男	167.6	56.2	89.8	35	33	55	49	112	7	235	31
男	157.1	50.5	85.8	33	29	48	57	70	7.4	205	29
男	165.4	61	85.2	34	31	45	54	76	8	237	22
男	168	60	91.1	40	31	55	52	76	7.5	225	23
男	165.9	49	89.7	37	32	62	56	87	7.8	240	26
男	170	61.5	91.2	36	31	48	55	68	8.2	212	28
男	168.7	57	92.3	40	42	50	62	102	6.9	240	37

図表7 整形したデータ



	身長	体重	座高	握力	上体起こし	長座体前屈	反復横跳び	シャトルラン	50m走	立ち幅跳び	ハンドボール投げ
身長	1.00										
体重	0.38	1.00									
座高	0.76	0.50	1.00								
握力	0.25	0.56	0.81	1.00							
上体起こし	0.07	0.09	-0.03	0.36	1.00						
長座体前屈	0.26	0.24	0.23	0.32	0.31	1.00					
反復横跳び	0.15	0.11	0.08	0.38	0.46	0.48	1.00				
シャトルラン	0.14	-0.08	0.03	-0.18	0.34	0.28	-0.37	1.00			
50m走	-0.21	-0.10	-0.21	-0.45	-0.33	-0.29	-0.54	-0.55	1.00		
立ち幅跳び	0.36	0.08	0.27	0.41	0.26	0.36	0.48	0.34	-0.67	1.00	
ハンドボール投げ	0.29	0.31	0.28	0.47	0.46	0.41	0.52	0.40	-0.49	0.42	1.00

政府統計の総合窓口 (e-Stat) からの時系列データ

政府統計の総合窓口 (e-Stat) からの時系列データ

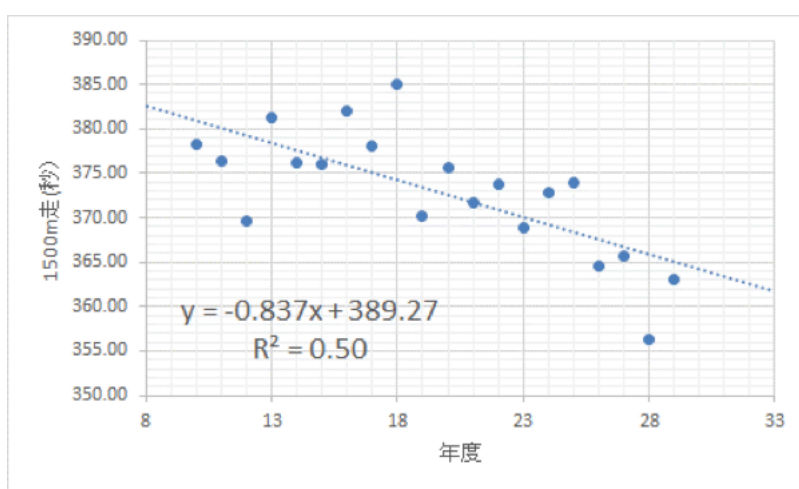
「科学の工具箱」からの個人の体力測定データ

図表8 散布図行列

上記の2種類のデータは項目も調査対象も類似しているが、一方は年度毎で集計された時系列データ（マクロな統計指標）で、もう一方は生徒一人ひとりを対象としたいわゆる個票とよばれるマイクロデータである。時系列データには、時間推移に伴うトレンド要因（例えば、その期間に全般的に体力や運動能力そのものが低下傾向にあるなど）が交絡因子として、項目に対して共通に関係する場合もあり、単純に散布図や相関係数から項目同士の直接的な関係と一概に判断するのは誤りである。一方で個人を対象としているデータは、項目間の関係はより直接的に解釈できる。時系列データに関しては時点間の変化の関係として傾向を捉える必要がある。

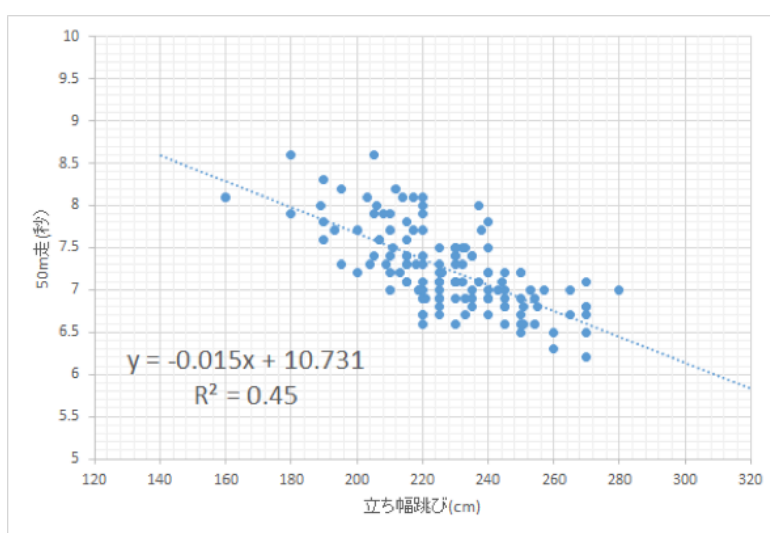
### ■手順3-2 単回帰分析による予測モデル構築

上記の2種類のデータそれぞれに、単回帰モデルを当てはめ、予測モデルを構築する。まず、時系列データに関しては、横軸Xに年度、縦軸Yに1500m走(秒)をとった散布図を作成する。グラフの点をクリックし、マウスの右メニューで「近似曲線の追加」を選択、「線形近似」を選びグラフに数式とR2を追加すると、図表9-Aが得られる。回帰係数は-0.837秒/年となり、1年あたり記録が平均で0.837秒縮む傾向にあることがわかる。この予測式に、x=30を代入すると、30年度の予測値を求めることができる。また、このモデルの寄与率は50%である。



図表 9-A 散布図に1500m走の記録の予測モデル(回帰直線)を入れた図

次に、個人の体力測定データに対して、立ち幅跳びの記録から50m(走)のタイムを予測する回帰モデルを同様の手順で求めると、図表9-Bとなる。回帰係数は、-0.015秒/cmなので、生徒間のばらつきから10cm立ち幅飛びの記録を高くすると、平均で50m走のタイムが0.15秒縮まる傾向にあることがわかる。また、このモデルの寄与率は45%である。

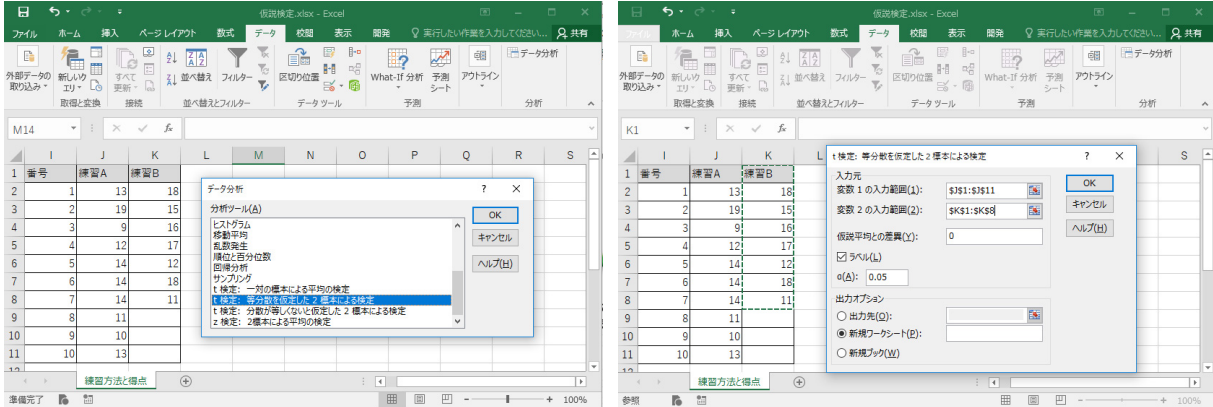


図表 9-B 立ち幅飛びの記録で50m走のタイムを予測するモデル(回帰直線)

## ＜演習 3＞

方法 A で練習した 10 名のグループと方法 B で練習した 7 名のグループの得点が右の表のように与えられている。このとき、2 つの練習方法に関して得点に差があるのか、母平均の差を仮説検定で調べてみましょう。

- (1) データは正規分布に従うことを仮定し、平均値の 2 標本の両側検定を選択、分散は 2 標本で共通で未知とする。
- (2) 帰無仮説  $H_0$  として「練習方法の効果に差はない (母平均は等しい)」とする。
- (3) 対立仮説  $H_1$  として「練習方法の効果に差はある (母平均は異なる)」とする。
- (4) 有意水準  $\alpha = 0.05$  (5%) と決める。
- (5) Excel の「データ分析」メニューから「t 検定: 等分散性を仮定した 2 標本による検定」を選択する。



(6) 出力から検定結果を読み取る (図表 10)

- ① 練習法 A, B 各々のグループの得点の標本平均
- ② 練習法 A, B 各々のグループの得点の標本分散
- ③ それぞれのグループのデータ数 (人数)
- ④ 2 つのグループの母分散 (共通の値) の推定値  $U_e$

$$U_e = \frac{(m-1)U_x + (n-1)U_y}{m+n-2}$$

ここで、 $m$  と  $n$  は各グループのデータ数、

$U_x, U_y$  は各グループの標本分散

- ⑤ 帰無仮説における 2 つのグループの母平均の差
- ⑥ 検定統計量  $t$  が従う  $t$  分布の自由度 ( $=m+n-2$ )
- ⑦ 検定統計量

$$t = \frac{|\bar{X} - \bar{Y}|}{\sqrt{U_e \left( \frac{1}{m} + \frac{1}{n} \right)}}$$

- ⑧ 片側検定の場合の有意確率 (p 値)
- ⑨ 有意水準  $\alpha$  (ここでは、5% に設定) の場合の片側検定での棄却限界値
- ⑩ 両側検定の場合の有意確率 (p 値)
- ⑪ 有意水準  $\alpha$  (ここでは、5% に設定) の場合の両側検定での棄却限界値

(7) 有意確率 P 値より結論を出す

この場合、両側検定と設定しているので、⑩より有意確率 (p 値) は 0.1026 で 0.05 を超えているので、帰無仮説を棄却することはできない。すなわち 2 つの練習方法の得点の母平均が異なるとは言えない。ここでは、分散が両群で等しいと仮定して計算している。

分散が両群で異なる場合は検定統計量  $t$  の計算は以下のようになる

$$t = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{U_x^2}{m} + \frac{U_y^2}{n}}}$$

	練習 A	練習 B	
1	t 検定: 等分散を仮定した 2 標本による検定		
2			
3			
4	平均	12.9	15.3
5	分散	7.66	7.90
6	観測数	10	7
7	プールされた分散	7.76	
8	仮説平均との差異	0	
9	自由度	15	
10	t	-1.738	
11	P(T<=t) 片側	0.0513	
12	t 境界値 片側	1.753	
13	P(T<=t) 両側	0.1026	
14	t 境界値 両側	2.131	

図表 10 t 検定の出力

## <参考文献・参考サイト>

- 「統計学基礎」日本統計学会編，東京書籍，2015年
- 「過去の気象データ・ダウンロード」気象庁，<https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
- 「科学の工具箱」<https://rika-net.com/contents/cp0530/contents/>
- 「政府統計の総合窓口（e-Stat）」<https://www.e-stat.go.jp/>

## ■ 学習活動と展開

### 【学習活動の目的】

- 名義尺度，順序尺度，間隔尺度，比例尺度などのデータの尺度水準の違いを理解する。
- 量的データについて，データ処理の流れ及びその評価について理解する。
- データを問題の発見・解決に活用するために，必要なデータの収集について，選択，判断する力を養う。
- データを多面的に精査して活用し，情報社会に主体的に参画しようとする態度を養う。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	アンケートや調査などで集まるデータを分類しよう。	アンケートや調査の項目を回答形式により，データの尺度について考える活動を行う。
展開 2	複数のデータ項目から関係性が強いデータを見つけよう。	相関係数や散布図行列を用いて数量データの関係性について調べる。
展開 3	体力測定 of 項目間の関係を調べてみよう。	複数のデータの中から 2 つの項目を選び，それらのデータの相関分析を行う。また，単回帰分析や統計的仮説検定を行う。

### 展開 1

問 い	アンケートや調査などで集められるデータの回答形式を分類しよう。
学習活動	<ul style="list-style-type: none"> <li>• アンケートや調査の項目を分類することにより，データの尺度について考える活動を行う。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>• 活動につまずいている生徒に，必要に応じてヒントを与え，生徒自らがデータの尺度を理解するよう促す。</li> </ul>





展開 2	
問 い	複数のデータ項目から関係性が強いデータを見つけよう。
学習活動	<ul style="list-style-type: none"> <li>相関係数や散布図行列を用いて量的データの関係性について調べる。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>相関係数を求められない生徒や散布図を作成できない生徒には、分析に必要な情報が得られるよう支援する。</li> <li>相関係数などの数値や散布図を読み取れない生徒に、データを読み取れるための観点を示すなどのヒントを与える。</li> </ul>



展開 3	
問 い	体力測定 of 項目間の関係を調べてみよう。
学習活動	<ul style="list-style-type: none"> <li>複数のデータ項目から、特定の2つの項目に着目して主な統計量を求める。</li> <li>箱ひげ図や統計的仮説検定により、2つのデータの分布について検討する。</li> <li>単回帰分析を行い、あるデータ項目を別データ項目で予測するモデルを見付ける。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>主な統計量が求められない生徒には、計算に用いる数式を確認するなどの支援をする。</li> <li>統計的仮説検定の行い方がわからない生徒には、どの段階で進めなくなっているかを確認して支援する。</li> <li>単回帰分析が理解できない生徒には、一次関数と関連させて理解できるようにする。</li> </ul>



まとめ	
まとめ	<ul style="list-style-type: none"> <li>量的データの分析のために必要な統計量の意味や求め方を確認して、それらを基に分析することの意義を確認する。</li> </ul>

## ■研修内容

### 【研修の目的】

- 質的データの種類やその扱い方について正しく理解することができる。
- テキストデータの扱いの基本について理解することができ、その基本的な収集や整理等の操作について理解できる。
- テキストデータの可視化についてその操作を行うことができ、それに関して自分の考えを基にした解釈や客観的な分析を行うことができる。
- 質的データの可視化や分析の基本について、理解することができ、その可能性についても説明できる。

### (1) 質的データの種類とその扱い

学習 22 において説明したように、データには尺度水準を基にすると、名義尺度、順序尺度、間隔尺度、比例尺度の 4 種類に分けて考えることができる。このうち、名義尺度と順序尺度をまとめて質的データまたはカテゴリ型データと呼ぶことがある。これらは量的データと異なり、一見すると数値で表されている場合などでも量的データと同様の演算や可視化を行うことができないことがあるので注意が必要である。質的データは、実際はテキスト（文字）などで表されるデータもあるが、それに数値が割り当てられている場合も多い。

回答形式①					
	1. あまり あてはまらない	2. やや あてはまらない	3. どちらとも いえない	4. やや あてはまる	5. よく あてはまる
文化祭に満足した。				○	

回答形式②  
あなたの満足度は、★★★★☆

図表 1 主観的な判断を回答させる場合

例えば、男性に 1、女性に 0 という数値が割り当てられているとき、その数値の平均値に意味はあるだろうか。確かに男女比を計算することが可能であり、プログラミングや表計算ソフトウェアにおいての集計では、よく利用される手法である。ところが、学年の項目に 1, 2, 3 などの数値が入力されている場合、これらの総和や平均値は何を意味するだろうか。発達段階を考慮した研究調査ならば、意味を持つ場合もあるかもしれない。アンケートなどを行う際などに仮に番号を付けたものなどについても同様のことが言える。それらの数値が順序しか意味を持たない場合、順序さえも意味を持たない場合などに分けて考えることにより、集計や平均値が意味を持つか否かが理解できる。右下のような主観的な判断を回答させる場合では、①と②では、回答者の意識が異なることなども理解させたい。一般的に、質的データに関しては、量的データと異なる処理と可視化が求められていることを理解したい。可視化に関しては、学習 24 で扱うこととする。

### (2) テキストデータの扱いについて

従来は、テキストなどの文字データに関しては、質的データとして使われたり、その表示内容として使用されたりすることが主であった。アンケートなどにおいても、選択肢やリッカート尺度などによる項目などの扱いに関しては、集計や分析がなされてきたが、自由記述の文章の分析をコンピュータで行うことができなかったために、人間による読解や手作業による分析や作業に頼らざるをえなかった。

近年、テキスト処理や自然言語に関するコンピュータによる処理技術が進み、文字認識や音声認識等の自然言語処理などと関連して、アンケート等の自由記述や小説などの長い文章に関する分析をすることも可能になった。

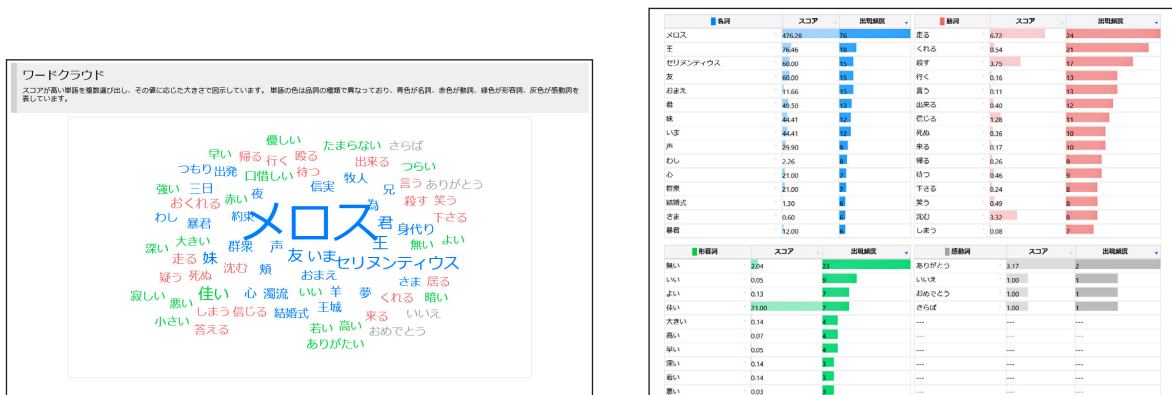
一般にコンピュータのプログラミング言語と区別して、人間の日常の会話や文章などに用いる日本語や英語などの言語を自然言語と呼ぶ。テキストデータの処理に関しては、英語と日本語では、その特性上大きく異なる。英語では、単語という区切りがあり、単語と単語の間には空白が空いているために単語を区切ることはそれほど難しくない。しかし、日本語の文章では、単語という区切りは、文章上からそのまま見てわかるわけではない。

例えば、有名な早口言葉である「すもももももものうち」などは、単語や品詞としてどのように分解すればよいのか、という処理が必要になる。また、英語でも日本語でも文法があり、品詞がある。また品詞によっては活用形があり、同じ品詞でも見かけ上は形が異なるものもあり、完全一致するものだけ検索することは難しい。

従来は人間が行っていたこのような処理をコンピュータで行うことで、様々な応用がなされている。アンケートの自由記述やお客様ががき・用紙の意見分析、小説等の作品・作者の傾向分析、Web上の記事の流行や内容の分析、TwitterやFacebook等のSNSなどの発言の分析やトレンドの抽出など、様々な分野で活用されている技術である。

### (3) テキストデータの可視化

ここでは日本語のテキストマイニングの基本について考えて、実際に実習をしてみよう。株式会社ユーザーローカルでは、様々なデータ分析のツールをWebベースで提供している。ソフトウェアのインストールなしにデータさえ用意すれば、手軽に利用できるのも、授業でも活用できる。今回は、(株)ユーザーローカルのユーザーローカルテキストマイニングツール (<https://textmining.userlocal.jp/>) を使ってみる。以下のデータはユーザーローカルテキストマイニングツールにサンプルとして掲載されている、太宰治『走れメロス』のデータを分析した結果である。



図表2 ワードクラウド

出典：ユーザーローカルテキストマイニングツール (<https://textmining.userlocal.jp/>)

ワードクラウド（左図）だけでなく、出現頻度グラフ（右図）、共起キーワードなどの様々な可視化が表示される。Twitterの発言内容等も可視化してみると、生徒の興味や関心を高めることができる。そして、その傾向などについて、生徒の主観的な解釈を含めて、考えていくことでテキストマイニングへの理解を深めることが可能である。

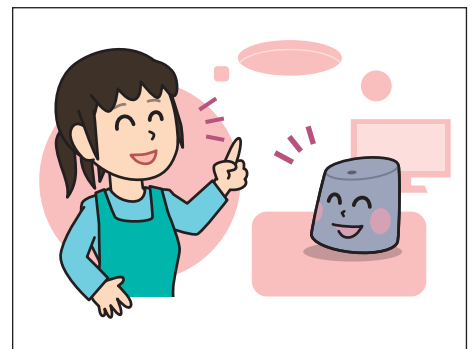
#### <演習1>

Twitterの発言内容やWebのテキストデータをユーザーローカルテキストマイニングツールを用いてテキストマイニングしてみましょう。出力された結果から、どのような傾向が考えられるか、考えてみましょう。

### (4) テキストの分析とその可能性

テキストマイニングを中心にテキストデータの分析についての話をしてきた。これらのテキストデータの分析にはどのような応用が考えられるだろうか。例えば、最初にあげた小説の作品分析などでは、文学作品の作者の言葉の使い方などについて分析でき、未発表の作品などについて、その作者が書いたものであるかの真贋を確かめることができる。テキストだけでなく、写真、絵画や彫刻などの作品に関しても機械学習や深層学習などの技術を用いて鑑定することが可能になる。また、後者のようなSNSの発言分析などでは、発言の真偽だけでなく、その発言者の傾向や変化なども分析することが可能になる。これまで、コンピュータによる分析というと数量的な分析が主であったものが、より人間の判断に近い分析ができるようになったということである。

また、これらの技術は、音声認識などにも活用されている。コンピュータやスマートフォン、スマートスピーカーに搭載されているAIアシスタント(Siri, Cortana, Alexaなど)も人間の音声をテキストに変換し、形態素解析を行い、意味解析を行い、その回答を返すように作られている。かつては単なるパターンマッチングであったものが、このようなテキストマイニング技術と機械学習などの技術を融合することにより、人間と同じような機能を持つ機器が登場してきている。



図表3 AIアシスタント

## <演習 2>

テキストマイニング技術によって、これからの身近な生活でどのような応用が考えられるか考えてみましょう。

### ◇発展

日本語のテキストマイニングを行うためには、文章を意味のある最小限の単位に分解する形態素解析をしなければならず、そのためのソフトウェアが必要である。有名なものとしては、「MeCab (めかぶ)」と「茶筌 (ちゃせん)」の2種類がある。テキストマイニングを行う際は、これらの形態素解析システムをプログラミング言語やアプリケーションソフトウェアから呼び出して使用する。

Windows ユーザーにはよく使われている「KH Coder」は「茶筌」を用いてテキストマイニングを行うアプリケーションソフトウェアである。ここでは、統計解析ソフトウェア R と MeCab を用いた簡単な例を紹介する。

R で MeCab を使うには

- (1) 統計解析ソフトウェア R のインストール
- (2) 形態素解析システム MeCab のインストール
- (3) R で MeCab を使うためのソフトウェア RMeCab のインストール
- (4) RMeCab の呼び出し

という手順が必要である。(1)～(3)は端末ごとにインストールが必要であり、(4)は MeCab を使用する際に宣言する必要がある。

(1)～(3)がインストールされている前提で、R を起動して形態素解析をしてみよう。

### プログラム

library(RMeCab)	説明：RMeCab を呼び出す
unlist(RMeCabc("隣の客はよく柿食う客だ"))	説明：形態素解析を行う

### 実行結果

名詞 助詞 名詞 助詞 形容詞 名詞 動詞 名詞 助動詞
"隣" "の" "客" "は" "良く" "柿" "食う" "客" "だ"

このように日本語の文章について意味を持つ最小限の単位に分割することが可能である。

次に、著作権フリーの作品を提供している「青空文庫」で芥川龍之介の「杜子春」の形態素解析をしてみよう。

手順は以下のとおりである。

- (1) パイプ関数の処理に必要な dplyr パッケージをインストールする。
- (2) パイプ関数の処理に必要な dplyr パッケージを読み込む。
- (3) R に「青空文庫」から作品を読み込むための Aozora 関数を読み込む。
- (4) Aozora 関数を使って、R に作品を読み込む。
- (5) 形態素解析を行う。

なお、Aozora 関数は RMeCab の作者である徳島大学の石田基広先生が提供している。

(1)がインストールされている前提で R を起動して形態素解析をしてみよう。

### プログラム

library(dplyr)	説明：dplyr パッケージの読み込み
source("http://rmecab.jp/R/Aozora.R")	説明：Aozora 関数の読み込み
x<-Aozora("https://www.aozora.gr.jp/cards/000879/files/43015_ruby_17393.zip")	説明：「杜子春」を x に読み込む
glove<-RMeCabFreq(x)	説明：x に対する操作

### 実行結果

# file = ./NORUBY/toshishun2.txt
# length = 1150

### プログラム

glove %>% head()	説明：x に対する操作
------------------	-------------

## 実行結果

```
# Term Info1 Info2 Freq
#1 あの フィラー * 1
#2 いいえ 感動詞 * 1
#3 おお 感動詞 * 2
#4 こら 感動詞 * 3
#5 さあ 感動詞 * 1
#6 はあ 感動詞 * 1
```

ここで、Term は単語（形態素）、Info1 は品詞の大分類、Info2 は品詞の細分類、Freq は形態素の出現頻度を表している。Subset 関数を使うと、選びたい品詞、除外したい品詞を指定することができる。以下では、大分類の中から名詞と動詞を選択し、細分類から数、非自立、接尾にあたる用語を除外している。

## プログラム

```
glove2 <- subset(glove,Info1 %in%("名詞","動詞"))
glove3 <- subset(glove2,!Info2 %in%("数","非自立","接尾"))
glove3 %>% head() 説明：選択及び除外したものに対する操作
```

## 実行結果

```
# Term Info1 Info2 Freq
#149 あく 動詞 自立 1
#150 ある 動詞 自立 18
#151 いう 動詞 自立 3
#152 見える 動詞 自立 1
#153 いる 動詞 自立 11
#154 かつまむ 動詞 自立 1
```

単語の出現頻度順にソートする際は、以下のようにする。

## プログラム

```
glove4 <- glove3[order(glove3$Freq,decreasing=T),] 説明：出現頻度順にソート
glove4 %>% head()
```

## 実行結果

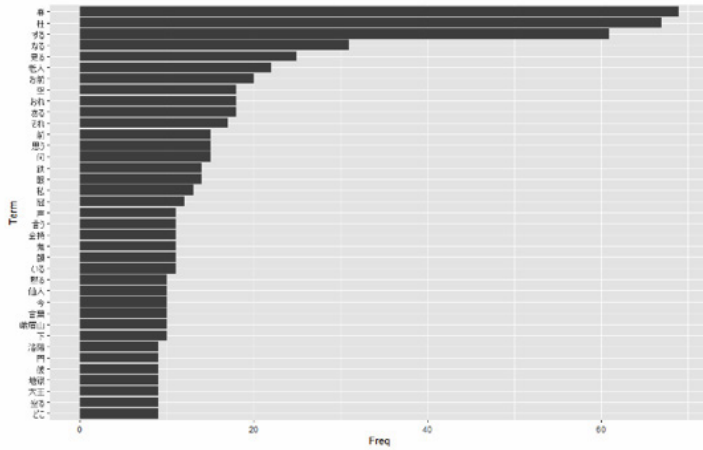
```
# Term Info1 Info2 Freq
#741 春 名詞 一般 69
#831 杜 名詞 一般 67
#163 する 動詞 自立 61
#179 なる 動詞 自立 31
#244 見る 動詞 自立 25
#929 老人 名詞 一般 22
```

テキストマイニングの手順がわかるように R で手順を追って説明してきた。さらに、R で出現頻度のグラフと出現頻に応じた文字の大きさで単語を表示するワードクラウドを作成することによって、テキストデータの可視化を行ってみよう。グラフを描くためには ggplot2、ワードクラウドを作成するためには wordcloud というパッケージのインストールが事前に必要である。

## プログラム

```
library(ggplot2)
glove4 %>%
  filter(Freq >=9) %>%
  mutate(Term=reorder(Term,Freq)) %>%
  ggplot(aes(Term,Freq))+
  geom_col()+
  theme_gray(base_family="IPAMincho")+
  coord_flip()library(ggplot2) 説明：ggplot2 の読み込み
説明：以下の条件で操作
説明：出現頻度 9 回以上
説明：頻度順にソート
説明：以下、グラフの描画

library(wordcloud)
wordcloud(glove4$Term,glove4$Freq,min.freq=4,color=brewer.pal(8,"Dark2"),family="IPAMincho")
```



図表4 ワードクラウド

ワードクラウドは「杜子春，地獄，峨眉山，老人」など作品のあらすじに関わる単語を表している。

<演習3>

青空文庫の作品をダウンロードして，統計ソフトウェア R を用いてテキストマイニングを行ってみましょう。

<参考文献・参考サイト>

- ・「ユーザーローカルテキストマイニングツール」 <https://textmining.userlocal.jp>
- ・「RMeCab」 <https://sites.google.com/site/rmecab/>
- ・「KHCoder」 <http://khcoder.net/>



## ■ 学習活動と展開

### 【学習活動の目的】

- 質的データの量的な処理の方法や可視化の方法について理解する。
- テキストデータの扱いの難しさ、英語と日本語の扱い方の違いなどを理解する。
- 簡単な形態素解析やストップワードについて理解する。
- テキストの頻度分析の可視化、タグクラウドなどを作成できるようにする。
- テキストマイニングを応用した音声認識、人工知能などについて、その仕組みを理解する。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	質的データと量的データの区別、質的データの処理について、理解しよう。	質的データを量的に扱った場合、どのような問題点が生じるかを考える。また、量的な処理を行いたいとき、アンケートなどでどのように収集すればよいのか、理解する。
展開 2	テキストマイニングとはどのような場面で活用できるのか考えよう。	テキストマイニング技術の可能性について話し合う。
展開 3	新見南吉の「手袋を買いに」と「ごんぎつね」では、どのような言葉の使い方の違いを解釈してみよう。	2つの小説やTwitterなどつぶやきなどからその傾向の違いについて比較する。

### 展開 1

問 い	質的データと量的データの区別、質的データの処理について、理解しよう。
学習活動	• 質的データ×質的データや質的データ×量的データの扱いについて理解する。
指導上の留意点	• 質的データと量的データの扱いの違いはどこにあるのか、同様に扱えない原因をもとに理解を深めるように促す。





展開 2	
問 い	テキストマイニングとはどのような場面で活用できるのか考えよう。
学習活動	<ul style="list-style-type: none"> <li>• テキストマイニングの技術について、その概略について理解する。</li> <li>• テキストマイニングが音声認識や文字認識などと合わせて、人の生活にどのような可能性を広げることができるか話し合う、その意見を共有する。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>• 奇抜な意見を尊重して、自由な発想を認めていく。</li> </ul>



展開 3	
問 い	新見南吉の「手袋を買いに」と「ごんぎつね」に見られる言葉の使い方の違いを解釈してみよう。
学習活動	<ul style="list-style-type: none"> <li>• テキストマイニングの基本的な操作について、プログラミング言語やアプリケーションを用いて身に付けさせる。</li> <li>• 二つの作品の共通点と相違点がワードクラウドなどに現れているかどうか、話合わせる。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>• グループごとに別な作品について、テキストマイニングを行い、相互に比較させてもよい。</li> </ul>



まとめ	
まとめ	<ul style="list-style-type: none"> <li>• テキストマイニングの技術について自ら実習を行うことで理解し、興味を持つことで、新たな可能性について考えることができるようにする。</li> </ul>

## ■ 研修内容

### 【研修の目的】

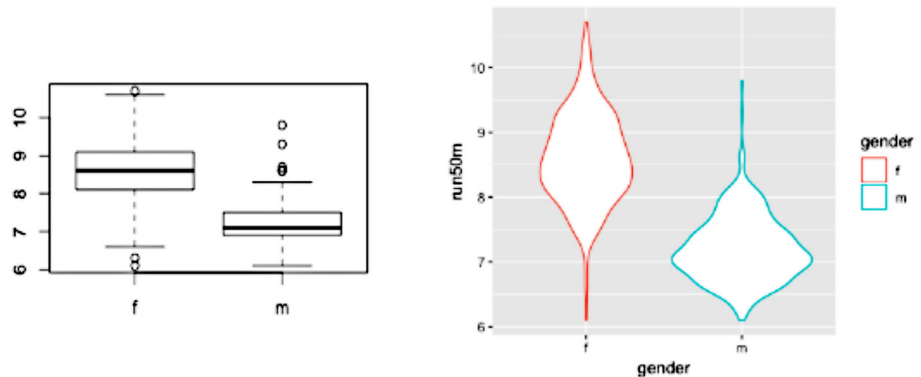
- 質的データの種類やその扱い方について正しく理解することができる。
- テキストデータの扱いの基本について理解することができ、その基本的な収集や整理等の操作について理解できる。
- テキストデータの可視化についてその操作を行うことができ、それに関して自分の考えを基にした解釈や客観的な分析を行うことができる。
- 質的データの可視化や分析の基本について、理解することができ、その可能性についても説明できる。

### (1) 質的データとその種類

名義尺度と順序尺度をまとめて質的データまたはカテゴリ型データと呼ぶことがある。これらは量的データと異なり、一見すると数値で表されている場合などでも量的データと同様の演算や可視化を行うことができないことがあるので注意が必要である。

例えば、クラスの生徒の男女別の身長について可視化するとしたら、どのような図を考えるだろうか。また、生徒の期末考査の英語と数学の成績を比較する際に素点ではなく5段階評定しか使うことができない場合、どのような可視化を行えばよいだろうか。

前者の例では、質的データ(名義尺度)×量的データ(比例尺度)と考えることができる。このとき、散布図などの量的データ×量的データの可視化の手法を使うことは望ましくない。適切でわかりやすい表現を行うことができないからである。総務省統計局の身体測定のサンプルデータから、箱ひげ図とヴァイオリンプロットを作成した。ヴァイオリンプロットは箱ひげ図では表現しきれない分布の様子を表すことができる可視化手法である。



図表1 箱ひげ図とヴァイオリンプロット

統計処理ソフトウェア R で、箱ひげ図は、以下のプログラムで描画することができる。

プログラム

```
boxplot(run50m~gender.data=gen_50)
```

ヴァイオリンプロットについては、以下のプログラムで描画することができる。

プログラム

```
library(ggplot2)
ggplot(data=gen_50, aes(x=gender, y=run50m, color=gender)) + geom_violin()
```

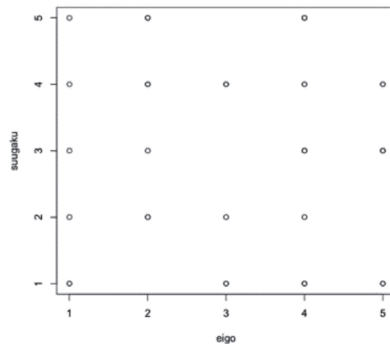
※ R のグラフ描画ライブラリ ggplot2 の導入方法

- ・ 初めて導入するとき、RStudio の Console から (要インターネット接続)  
install.packages("ggplot2")
- ・ 導入したものを使用するとき  
library(ggplot2)

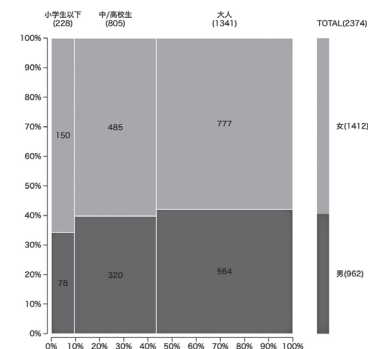
後者の例はどうなるだろうか。質的データ(順序尺度)×質的データ(順序尺度)である。ここでは、仮にランダムに40人の5段階のデータを作成した。分割表(クロス集計表とも呼ばれている)は、可視化手法の一つであり例えば図表2のように表すことができる。縦が英語、横が数学である。英語が2で数学が5の生徒は3人いると読み取ることができる。これを散布図で表すこともあるが、図表3のようになり、よく表現できていない。

		英語				
		1	2	3	4	5
数学	1	2	0	2	3	2
	2	1	2	1	1	0
	3	1	1	0	3	3
	4	1	3	3	2	2
	5	1	3	0	3	0

図表2 分割表



図表3 散布図



図表4 マリメッコチャートの例

このような、質的データ×質的データにおいては、分割表なども用いることが多いが、端的に可視化するときには、モザイク図や図表4のマリメッコチャートなども用いられる。

### <演習1>

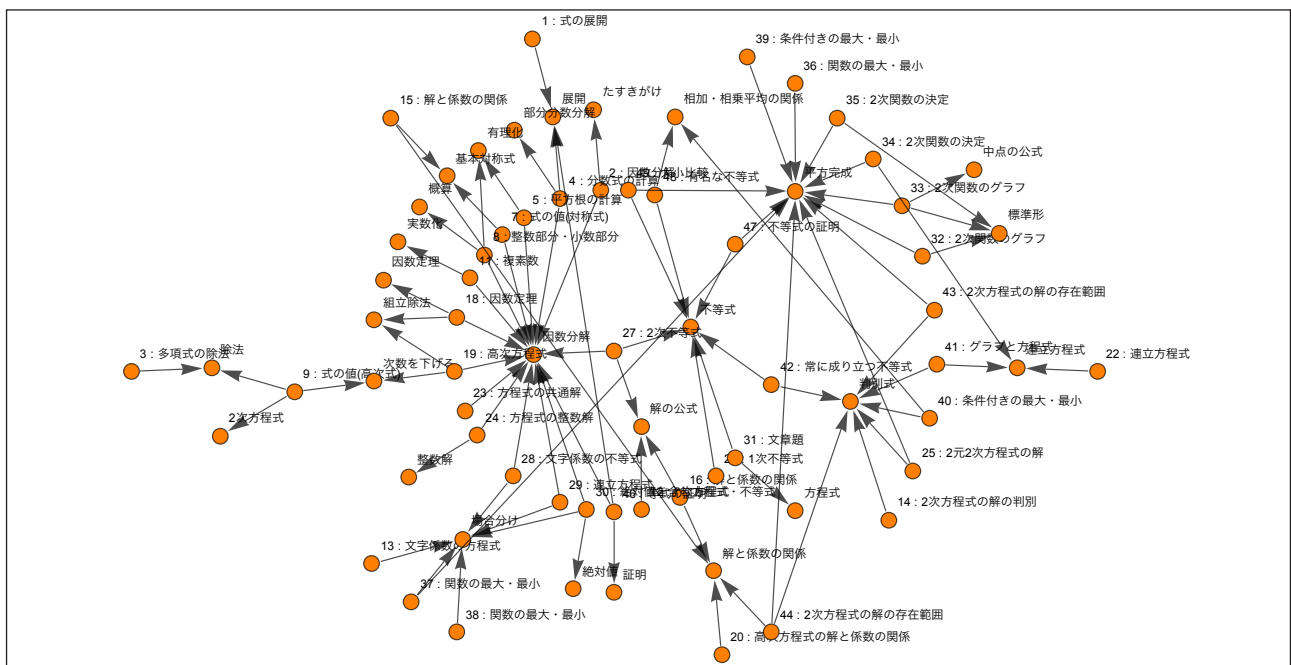
- (1) 図表3の散布図は、どのような点に問題があるのでしょうか。話し合ってみましょう。
- (2) モザイク図について調べ、統計ソフトウェアまたは表計算ソフトウェアなどを用いて、質的データをモザイク図で表現してみましょう。

## (2) データの分析と可視化

データは形式や分析の目的に合った可視化ができる。

前節では、質的データの可視化についての説明をした。数学のデータの分析で学ぶ可視化については、量的データの可視化が主であるために、質的データでも同様の可視化手法を行ってしまうことも多く見られるが、質的データにおいては、それに対応した可視化が存在することを意識させたい。

さて、表計算やリレーショナルデータベースで表現されている矩形データの可視化については後述するが、別の形式データの可視化についても考えてみよう。例えば、数学の問題集の問題と解法の関係などは矩形データで表すよりも関係図で示すことが適切である場合が多い。



図表5 関係図の例(問題と解法)

参考:「チャート式シリーズ 入試必携168」数研出版

さて、矩形データに関しては、表計算ソフトウェアのグラフ機能を使って表すことも一つの方法ではあるが、先にあげた統計解析ソフトウェア R など で記述して、可視化することも可能である。また可視化専用のソフトウェアも多く存在する。Tableau (タブロー) Desktop は、Tableau 社が提供している可視化ソフトウェアである。教育機関での授業での利用や学生が使用する場合は、無償のライセンスを提供する Tableau Academic プログラムがある。Excel, CSV など様々な形式のデータに対応しており、データを読み込んだ後は、ドラッグ & ドロップやダブルクリックなどの操作で手軽に可視化することが可能である。BI (Business Intelligence) ツールとして提供されているツールであり、企業の経営意志決定に使える本格的なツールとしても有名なツールである。チュートリアルビデオなどもあり、使い方を自学自習で学ぶことができ、Excel など で、データの加工やクリーニングを行い、Tableau Desktop で可視化するというような使い方もできる。また、Excel などのデータを読み込んだ時点で、データの項目を市町村別就業人口などの分析の切り口であるディメンションと市町村ごとの就業人口などの指標であるメジャーに分けてくれる、これは少し異なる部分もあるが、質的データと量的データの判別をしてくれるとして考えてもよいだろう。また可視化の種類を選択する際にもメジャーを選択するのかディメンションを選択するのかのガイドも表示してくれるので、わかりやすい。

通常の矩形データでも Tableau Desktop で使用することは可能であるが、更に様々な情報を付け加えると更に表現力が広がる。郵便番号や都道府県、市町村名などを含めることで地図上にグラフを表示することなども簡単に行うことができる。

## <演習 2>

有向グラフで表すことができる例を探し、それを有向グラフで可視化できるツールにより、可視化してみましょう。また、作成するツールがない場合は、紙に書いて表現してみましょう。

### (3) データの可視化と問題発見

データの可視化は何のために行うのだろうか。可視化することによって、問題を発見することができ、それに対して詳細な分析や解釈、解決策を考えることができるからである。ここでは、統計解析ソフトウェア R の ggplot2 というパッケージに含まれている diamonds サンプルデータを用いて説明しよう。このデータは、実際に米国で流通したダイヤモンドのカラット、カット、透明度、大きさ、価格などを含んだ数万件の大きなデータである。

※ R のデータフレーム操作ライブラリ dplyr の導入方法 (再掲)

- ・ 初めて導入するとき、RStudio の Console から (要インターネット接続)  
install.packages("dplyr")
- ・ 導入したものを使用时  
library(dplyr)

※ R のグラフ描画ライブラリ ggplot2 の導入方法 (再掲)

- ・ 初めて導入するとき、RStudio の Console から (要インターネット接続)  
install.packages("ggplot2")
- ・ 導入したものを使用时  
library(ggplot2)

#### プログラム

```
> library(dplyr)
> library(ggplot2)
> diamonds
```

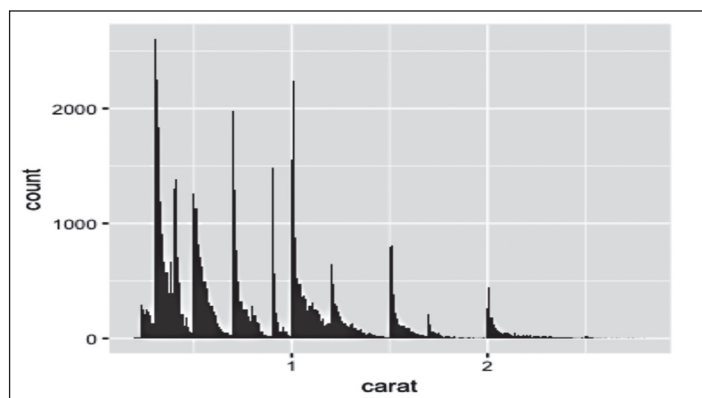
#### 実行結果

```
# A tibble: 53,940 x 10
  carat cut    color clarity depth table price    x    y    z
  <dbl> <or> <ord> <ord> <dbl> <dbl> <int> <dbl> <dbl> <dbl>
1  0.23 Ide... E   SI2   61.5   55   326   3.95  3.98  2.43
2  0.21 Pre... E   SI1   59.8   61   326   3.89  3.84  2.31
3  0.23 Good  E   VS1   56.9   65   327   4.05  4.07  2.31
4  0.290 Pre... I   VS2   62.4   58   334   4.2   4.23  2.63
5  0.31 Good  J   SI2   63.3   58   335   4.34  4.35  2.75
6  0.24 Ver... J   VVS2  62.8   57   336   3.94  3.96  2.48
7  0.24 Ver... I   VVS1  62.3   57   336   3.95  3.98  2.47
8  0.26 Ver... H   SI1   61.9   55   337   4.07  4.11  2.53
9  0.22 Fair  E   VS2   65.1   61   337   3.87  3.78  2.49
10 0.23 Ver... H   VS1   59.4   61   338   4     4.05  2.39
# ... with 53,930 more rows
```

このデータをもとにダイヤモンドのカラット数を 0.01 刻みでグラフを描くと面白いことが分かる。

## プログラム

```
smaller <- diamonds %>% filter(carat < 3)
ggplot(data = smaller, mapping = aes(x = carat)) + geom_histogram(binwidth = 0.01)
```



図表6 グラフの例 (ダイヤモンドのカラット数)

転載:「Rではじめるデータサイエンス」P78 グラフ ((株) オライリー・ジャパン)

このグラフは、なぜか右側に裾を引くような形を繰り返している。これも不思議な現象であるが、このグラフに関して、話し合い各自の解釈を考えることも大切な実習である。グラフの表示結果は、客観的であり、みな同じものが表示されるかもしれないが、それぞれの読み取り方(解釈)を重視することもデータの分析では大切なスキルの一つである。

次のプログラムは、横軸にカラット数、縦軸に価格として散布図を描くためのものである。このグラフから問題を発見することも可能である(グラフは省略する)。

## プログラム

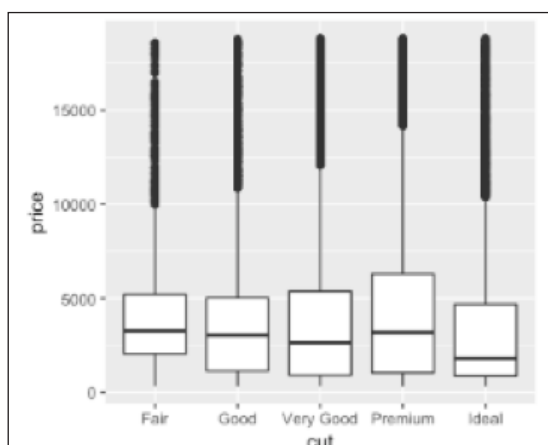
```
library(ggplot2)
ggplot(data = diamonds) + geom_point(mapping = aes(x = carat, y = price))
```

数学科の授業の過程で数万個のデータを描画することは考えにくい。データが多いからこそその問題も見つけることができる。点が重なってしまっていて、どこに点が集中しているのかが分からなくなる場合もあるが、現実社会で扱うデータでは、このようなことはよく見られることである。

次の問題も問題提起である。下の箱ひげ図は、横軸にダイヤモンドのカット(cut)、縦軸に価格(price)を設定して描いたものである。横軸は右に行くほどカットの質が上がる。

## プログラム

```
ggplot(diamonds, aes(cut, price)) + geom_boxplot()
```



図表7 箱ひげ図の例 (ダイヤモンドのカットの質と価格)

転載:「Rではじめるデータサイエンス」P85 箱ひげ図 ((株) オライリー・ジャパン)

これは不思議な現象に見えるだろう。カットの質が上がるほど、価格が下がっている。色や透明度の高さなどでも同様の逆転現象が起こる。ヒントは「交絡因子」である。交絡因子とは、対象としている2変数と相互に相関の高い隠れた変数（因子）のことである。

### <演習3>

なぜ、このような逆転現象が起こっているのでしょうか。考えてみましょう。また、どのようなグラフを描けば、この現象について納得できる説明ができるのでしょうか。データをもとに実際にグラフを描いてみましょう。

可視化はデータの分析の一部でもあり、問題の発見や現象の説明に活用することができる。また、予測を立てることに活用することができる。

### <参考文献・参考サイト>

- ・「Rではじめるデータサイエンス」Hadley Wickham, Garrett Grolemund 著, オライリージャパン, 2017
- ・「チャート式シリーズ 入試必携 168」数研出版
- ・「Tableau Desktop」<https://www.tableau.com/ja-jp/products/desktop>



## ■ 学習活動と展開

### 【学習活動の目的】

- データの形式に応じた可視化手法や分析目的に応じた可視化手法を選択できるようになる。
- 不適切や誤解を招く可視化について理解する。
- データの様々な可視化によって、問題発見、解決ができることを理解する。
- 問題解決のためにどのような可視化を選択すべきか判断できるようになる。

### ○ 学習活動とそれを促す問い

	問 い	学習活動
展開 1	箱ひげ図はどのような時に使うべきか、使い方を理解しよう。また、箱ひげ図とヒストグラムを区別しよう。	可視化手法に関して、その目的や特性について理解させる。
展開 2	表されたグラフから問題を発見しよう。考えられる問題点について自分の解釈も含めて説明してみよう。	グラフから見える特徴や外れ値、傾向などを読み取り、より良い可視化手法を用いて改善や問題の明確化を行う方法について学習する。
展開 3	立体グラフや円グラフが誤解を招く表現であるとされるのはなぜか、調べてみよう。	不適切なグラフや誤解を招く表現について理解し、自己の可視化表現に関して注意すべき点について整理する。

展開 1	
問 い	箱ひげ図はどのような時に使うべきか、使い方を理解しよう。また、箱ひげ図とヒストグラムについて理解しよう。
学習活動	<ul style="list-style-type: none"> <li>• データの種類と可視化の目的によって、可視化の手法は異なることを理解する。</li> <li>• 特に質的データの可視化に関しては、様々な表現があることを理解する。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>• 実際のデータをもとにして、グラフを書かせることが望ましい。</li> <li>• また、可視化専用のソフトウェアの活用なども検討する。</li> <li>• データの種類によって、様々な可視化手法があることを提示する。</li> </ul>





展開 2	
問 い	表されたグラフから問題を発見しよう。考えられる問題点について自分の解釈も含めて説明してみよう。
学習活動	<ul style="list-style-type: none"> <li>・可視化から問題の発見に繋がることを理解する。</li> <li>・データが多くなると可視化する上で留意することが出てくることに注意する。</li> <li>・交絡因子の存在について理解する。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>・可視化されたデータから、何が問題であるのか、問題は1つとは限らないことを理解させる。</li> <li>・また、可視化によって問題の解決に繋がることも促したい。</li> </ul>



展開 3	
問 い	立体グラフや円グラフが誤解を招く表現であるとされるのはなぜか、調べてみよう。
学習活動	<ul style="list-style-type: none"> <li>・実際の誤った可視化の使い方などを例にその表現について理解する。</li> </ul>
指導上の留意点	<ul style="list-style-type: none"> <li>・ニュースなどに用いられている可視化に関して、批判的な態度で評価することができるようにする。</li> <li>・また、誤った例をできるだけ多く提示する。</li> </ul>



まとめ	
まとめ	<ul style="list-style-type: none"> <li>・可視化が問題発見や解決の手法となっていることを理解する。</li> </ul>

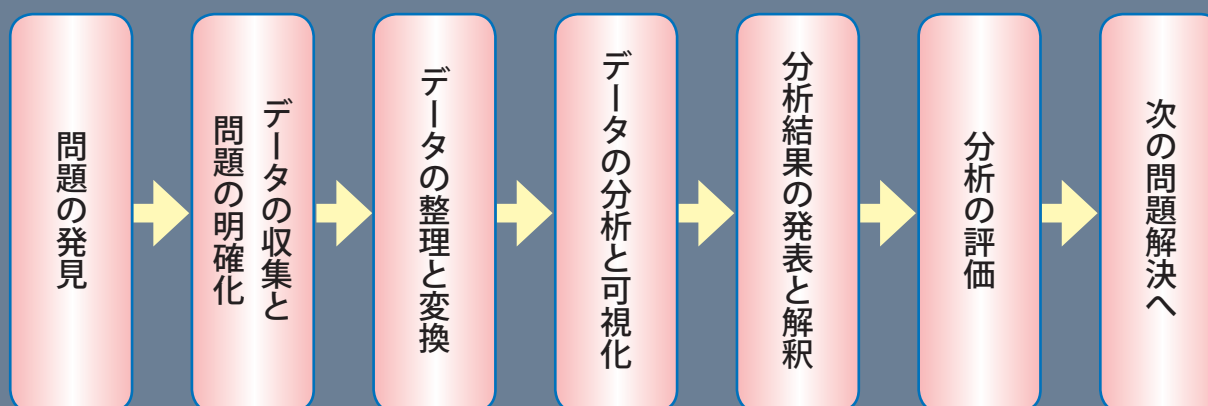
# ◆全体を通じた学習活動の進め方

## 【全体を通じた学習活動の目的】

- ・情報通信ネットワークの管理、運用ができ、データベースや Web 上のテキストデータ、オープンデータ等を可視化、分析する力を育成する。

## 【全体を通じた学習活動の流れ】

データの分析に関する実習については、以下のような手順で行うことが望ましい。



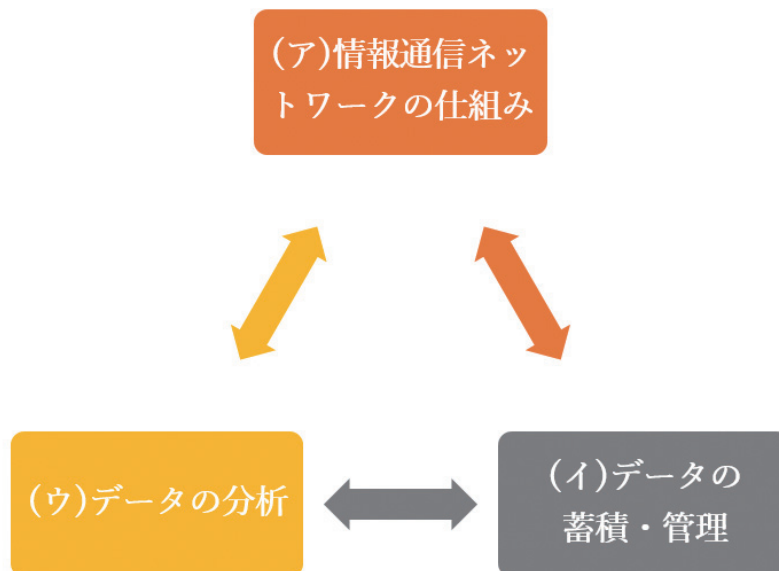
実習に関しては、以下のような項目と内容に留意して行うことが望ましい。

①情報通信ネットワークの仕組みとデータの蓄積・管理	項目(ア)「情報通信ネットワークの仕組み」と(イ)「データの蓄積・管理」に関しては、情報通信ネットワークを活用した情報システムについて考え、そこでデータを蓄積する仕組みやそのセキュリティに関して調べ、安全な情報システムとは何か、考察することなどが考えられる。また、情報通信ネットワーク上のデータは、リレーショナルデータベースのような形で管理されているものばかりではなく、様々な形式で管理されていることにも理解が必要である。
②データの蓄積・管理とデータの分析	項目(イ)「情報の蓄積・管理」と(ウ)「情報の分析」に関しては、情報通信ネットワーク上のデータがリレーショナルデータベースの形式のものやそれ以外の形式のものも取り込み、分析や可視化しやすいように整理、変換する実習などが考えられる。学習20のオープンデータの活用や学習25などで取り組んだテキストマイニングに関しても、積極的に Web 上のテキストや公開されているテキストデータで興味あるものを可視化や分析が可能なように整理し、処理することによって相互の項目の関連性を深く理解することができる。Web 上のデータなどは必ずしも処理しやすい形式にはなっていないが、それを質的/量的データの区別を意識して、整理することによって、可視化や分析の手法が決定できることを理解させたい。

<p>③ 情報通信ネットワークとデータの分析</p>	<p>項目(ア)「情報通信ネットワークの仕組み」と(ウ)「情報の分析」に関しては、データとパケット、プロトコルなどの関連性を意識し、どのようなデータであれば通信するのに向いているか、またビッグデータを即時に通信するには、どのような手段が考えられるか、個人情報を含むデータなどを通信するにはどのような方法が考えられるか、などについてデータの形式と通信についての関連性に関して考えることができる。またネットワークの接続を表すグラフとデータの形式などに関しても、ルータのルーティングテーブルやフィルタなどに関連した学習を考えてもよいだろう。</p>
<p>④ データの分析に関する総合演習</p>	<p>データの分析に関する演習については、収集、整理、分析・可視化、評価の一連の流れを実施することが望ましい。使用するデータはきれいに整形したデータではなく、多少加工が必要なデータを用いて、実際のデータがどのようなものであるか、理解させることも必要である。データの整理には、分析や可視化の目的に合わせた変換が必要なことや、質的/量的データの分析や可視化の方法の選択も考慮することが必要になる。学習内容として挙げたオープンデータなどを積極的に活用し、数十件程度のデータではなく、少なくとも数千件あり単にデータを眺めただけでは傾向が読み取れないものに関して、問題や結果を発見するような実習を設定することが望ましい。分析や可視化に関しては、(3)のプログラミングと関連して、プログラミング言語を活用してもよいが、(3)で学んだ言語が必ずしもデータ分析に適した言語とは言えない場合もあるので、データ分析に適したプログラミング言語を選択することも必要である。</p>

**【全体を通じた学習活動を行ううえでの注意点】**

この単元は、異なる3つの項目からなっており、3つすべてを通じた学習活動に関しては、実施することが難しいかもしれない。2つの項目を有機的に接続して、総合的に考えていく実習などを考えることが必要である。



平成30年度文部科学省委託  
「次世代の教育情報化推進事業」  
(高等学校情報科担当教員の指導力向上に関する調査研究)

【有識者名簿】

(1) 検討委員

- 主査 中川 一史 …………… (放送大学 教授)  
委員 能城 茂雄 …………… (東京都立三鷹中等教育学校 主幹教諭)  
委員 小澤 美紀 …………… (神奈川県立総合教育センター 指導主事)  
委員 石田 淳一 …………… (株式会社アールジェイ 代表取締役)

(2) WG委員

- WG 1 小原 格 …………… (東京都立町田高等学校 指導教諭)  
WG 1 小松 一智 …………… (東京都立石神井高等学校 指導教諭)  
WG 2 能城 茂雄 …………… (東京都立三鷹中等教育学校 主幹教諭)  
WG 2 上平 崇仁 …………… (専修大学 教授)  
WG 2 梅沢 崇 …………… (東京都立新宿山吹高等学校 主幹教諭)  
WG 3 兼宗 進 …………… (大阪電気通信大学 教授)  
WG 3 天良 和男 …………… (東京学芸大学 特任教授)  
WG 3 鎌田 高德 …………… (神奈川県立茅ヶ崎西浜高等学校 教諭)  
WG 3 三井 栄慶 …………… (神奈川県立横浜翠嵐高等学校 教諭)  
WG 3 日々野清高 …………… (iEdTech 代表)  
WG 4 大橋 真也 …………… (千葉県立千葉中学校・千葉高等学校 教諭)  
WG 4 春日井 優 …………… (埼玉県立川越南高等学校 教諭)  
WG 4 渡辺美智子 …………… (慶應義塾大学大学院 教授)

## 【学会等連絡先】

研修講師を依頼する際などの窓口として御利用下さい。

(学会名五十音順)

- (1) **情報処理学会** 担当：萩原 恵子 03-3518-8372 sig@ipsj.or.jp  
〒101-0062 東京都千代田区神田駿河台 1-5 化学会館 4F  
一般社団法人 情報処理学会研究部門
- (2) **日本教育工学会** 担当：長尾 正子 03-5549-2263 office@jset.gr.jp  
〒107-0052 東京都港区赤坂 1-9-13 三会堂ビル 8階  
日本教育工学会事務局
- (3) **日本教育情報学会** 担当：若杉 祥太 090-1026-1413 jsei@ashiya-u.ac.jp  
〒659-8511 兵庫県芦屋市六麓荘町 13-22 芦屋大学技術研究棟若杉研究室  
日本教育情報学会運営本部事務局
- (4) **(一社)日本産業技術教育学会** 担当：岡田 和美 075-415-3661 jste@nacoss.com  
〒602-8048 京都市上京区下立売通小川東入西大路町 146  
中西印刷株式会社 学会部内  
(一社)日本産業技術教育学会事務局
- (5) **日本情報科教育学会** 担当：阿濱 茂樹 083-933-5392 ahama@yamaguchi-u.ac.jp  
〒753-8511 山口県山口市吉田 1677-1 山口大学教育学部阿濱研究室  
日本情報科教育学会 教員養成・研修委員会
- (6) **(一社)日本統計学会  
統計教育委員会** 事務局：sesjss@stat.k-junshin.ac.jp
- (7) **(一社)日本品質管理学会  
TQE 特別委員会** 事務局：apply@jsqc.org

## 高等学校情報科「情報Ⅰ」教員研修用教材

---

---

◇発行日◇

平成31年3月29日

◇発行◇

文部科学省

◇制作◇

エヌ・ティ・ティラーニングシステムズ株式会社

---

---