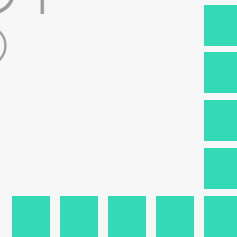


データサイエンスマンダラート

学籍番号: 2122041 氏名: 富原光流 所属: グループ 1
活動時間: スライド作成 (15時間) 音声録音 (1時間)
方法: Google slide, Online Audio Converter
Online Audio Joiner, ゆくも



データサイエンスマンダラートとは



データサイエンスマンダラートとは、
統計学、人工知能学、データ工学、情報学の確定領域と
グループ選択領域の**数学とプログラミング**からなる講義資料であると定義する。

<u>データベース</u> (富原)	<u>データマイニング</u> (富原)	<u>ビッグデータ</u> (青木)				<u>記述統計学</u> (ミン)	<u>統計応用</u> (浅井)	<u>推計統計学</u> (浅井)
<u>データウェアハウス</u> (富原)	データ工学	<u>OLAP</u> (浅井)				<u>疫学統計</u> (浅井)	統計学	<u>社会統計学</u> (ミン)
<u>マルチメディア</u> (ガワンデ)	<u>ロボティクス</u> (浅井)	<u>クラウド コンピューティング</u> (ミン)				<u>要約統計学</u> (青木)	<u>多変量解析</u> (青木)	<u>ベイズ統計学</u> (青木)
<u>エラー</u> (ガワンデ)		<u>オブジェクト</u> (富原)	<u>データ工学</u>		統計学	<u>期待値</u> (青木)		<u>標準偏差</u> (浅井)
<u>ソースコード</u> (ミン)	<u>プログラミング</u>	<u>関数</u> (浅井)	<u>プログラミング</u>	DS学	数学	<u>分散</u> (富原)	数学	<u>相関関係</u> (ミン)
<u>オープンソース</u> (青木)		<u>メソッド</u> (富原)	<u>人工知能学</u>		情報学	<u>ヒストグラム</u> (浅井)		<u>平均値</u> (ガワンデ)
<u>チューリングテスト</u> (ガワンデ)	<u>機械学習</u> (ミン)	<u>ディープラーニング</u> (ミン)				<u>アルゴリズム</u> (青木)	<u>ソフトウェア</u> (ミン)	<u>セキュリティ</u> (富原)
<u>音声処理</u> (青木)	人工知能学	<u>自然言語処理</u> (富原)				<u>ヒューマンエラー</u> (青木)	情報学	<u>サイバー犯罪</u> (浅井)
<u>画像処理</u> (青木)	<u>シンギュラリティ</u> (浅井)	<u>ニューラル ネットワーク</u> (浅井)				<u>情報技術</u> (ガワンデ)	<u>サーバー</u> (浅井)	<u>フィンテック</u> (浅井)

統計学



[マンダラートマップへ](#)

[次の領域へ](#)

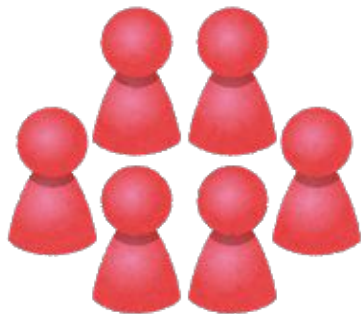
推計統計学とは？



限られた標本から調査したい

母集団全体の特徴を推測するという学問

母集団



標本



データを発生させる構造

データ

疫学とは

地域や特定の人間集団の中で出現する健康事象(病気など)の発生頻度や要因を明らかにする科学研究。



疫学統計とは

結果の頻度や分布を調べることにより、原因と結果に関する**特性を調べたり、原因の仮説を立てる**学問。

多変量解析とは…

観測値が複数の値からなるデータ(多変量データ)を統計的に扱う手法

多変量解析でできること

- アンケートの結果から商品の強み・弱みを知りたいとき
- 身体測定 of データから病気になる確率を知りたいとき
- 既存店舗の売上や顧客数などのデータから、新店舗の将来の売上を予測したいとき

多変量解析の目的は「予測」と「要約」の2つがあります。

目的	内容	例
予測	未来の出来事を予測値から測定。	共通テストにおいて各教科の点数と志望校の偏差値を把握していれば、合格率が予想できます。
要約	複数の測定値をまとめて、ある要素について解析。	英語、数学、理科の3教科のテストから理系としての能力を解析できます。

参考:<https://udemy.benesse.co.jp/data-science/data-analysis/multivariate-analysis.html>

ベイズ統計学とは…

データが不十分でもある事態が発生する確率(事前確率)を最初に設定する。
更に情報が得られたらある事態が発生する確率(事後確率)を更新していき、
本来起こるであろう事象の確率(主観確率)を導き出します。



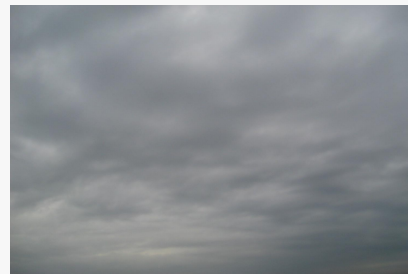
今日の天気は曇り？ 晴れ？
なんとなく晴れる確率は
50%だろう。

事前確率

確率の更新

時間がかかったらもっと曇りになっ
た。晴れる確率は
10%だろう。

事後確率



参考・引用:<https://ai-trend.jp/basic-study/basic/bayesian-statistics/>

大量のデータを得られたら何をする???



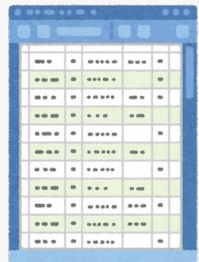
調べたデータ量が膨大に増えてしまったら、データを **一つ一つ確認** することは **不可能**



では、データを要約しよう!!

要約統計学とは

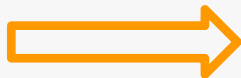
できるだけ多くの情報を簡単に伝えるために、要約統計を使用して**一連の観測を要約すること**



日	時間	気温	湿度	風速	降水
2023-10-01	00:00	15.0	65.0	0.0	0.0
2023-10-01	01:00	14.0	68.0	0.0	0.0
2023-10-01	02:00	13.0	70.0	0.0	0.0
2023-10-01	03:00	12.0	72.0	0.0	0.0
2023-10-01	04:00	11.0	75.0	0.0	0.0
2023-10-01	05:00	10.0	78.0	0.0	0.0
2023-10-01	06:00	9.0	80.0	0.0	0.0
2023-10-01	07:00	8.0	82.0	0.0	0.0
2023-10-01	08:00	7.0	85.0	0.0	0.0
2023-10-01	09:00	6.0	88.0	0.0	0.0
2023-10-01	10:00	5.0	90.0	0.0	0.0
2023-10-01	11:00	4.0	92.0	0.0	0.0
2023-10-01	12:00	3.0	95.0	0.0	0.0
2023-10-01	13:00	2.0	98.0	0.0	0.0
2023-10-01	14:00	1.0	100.0	0.0	0.0
2023-10-01	15:00	0.0	100.0	0.0	0.0
2023-10-01	16:00	1.0	98.0	0.0	0.0
2023-10-01	17:00	2.0	95.0	0.0	0.0
2023-10-01	18:00	3.0	90.0	0.0	0.0
2023-10-01	19:00	4.0	85.0	0.0	0.0
2023-10-01	20:00	5.0	80.0	0.0	0.0
2023-10-01	21:00	6.0	75.0	0.0	0.0
2023-10-01	22:00	7.0	70.0	0.0	0.0
2023-10-01	23:00	8.0	65.0	0.0	0.0

収集した大量
のデータ

画像引用: <https://www.irasutoya.com/>

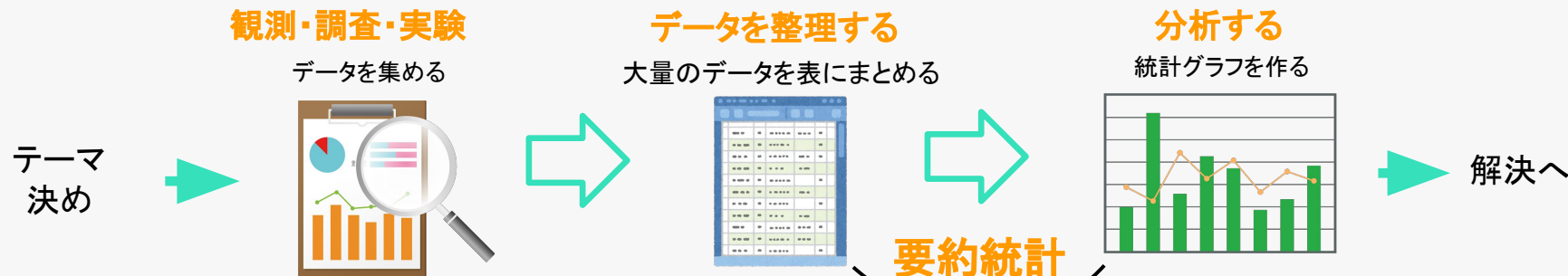


例えば、average関数を使用し
て平均値を出した。



出した平均値が要約統計

要約統計学



参考:<https://best-biostatistics.com/summary/sum-stat.html#i-2>

画像引用:<https://www.irasutoya.com/>

要約統計代表的なもの

代表値

全体の中心的な位置を
表現する数値。

- 平均値 (average)
- 中央値 (median)
- 最頻値 (mode)

散布度

全体の散らばりを
表現する数値

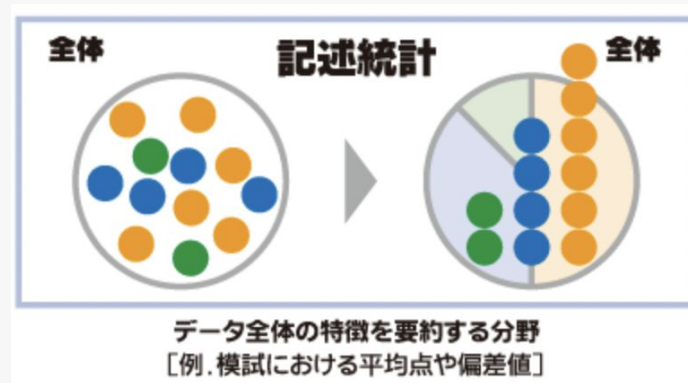
- 標準偏差
- 分散
- 平均偏差

記述統計学(descriptive statistics)



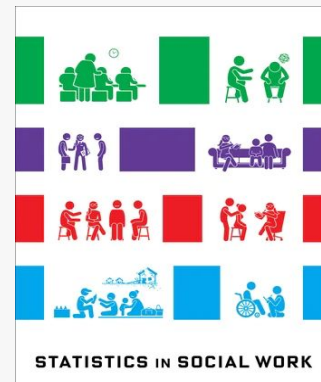
記述統計学とは観測して得られた各データを整理したり要約したりする方法を学ぶ学問。

表やグラフを作成したり、平均値や標準偏差を求めることなどによって、手持ちのデータの特徴をより簡単にわかりやすく表現する。



社会統計学(Social Statistics)

社会統計学とは社会学などの **社会科学**に応用する統計学。
簡単に言うと、**社会調査データ**を分析する統計学。



社会統計学、3つの目的

1. グループ、組織が利用できるサービスの品質を評価する。
2. ある環境や特別な状況における人々の行動を分析する。
3. 統計的サンプリング(標本調査)を通じて人々の欲求を推測する。

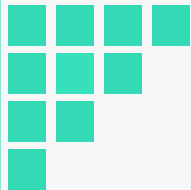
統計応用とは？



数学の一分野であり、数学的モデルやコンピュータを利用して
理論的，実践的課題を解決するための方法が検討される学問

	A	B
1	timestamp	AccX
2	1620369279104	-1.87679718
3	1620369279120	-1.976190491
4	1620369279184	-3.141367493
5	1620369279252	-1.558259583
6	1620369279318	0.3478765869
7	1620369279384	-0.605640564
8	1620369279451	-1.841320953
9	1620369279517	-17.42840744
10	1620369279584	-45.03115723
11	1620369279652	3.055446167
12	1620369279717	-69.33985703
13	1620369279784	-64.20134262

数学的な処理



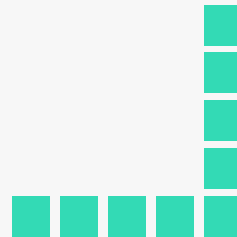
データ工学



[前の領域へ](#)

[マンダラートマップへ](#)

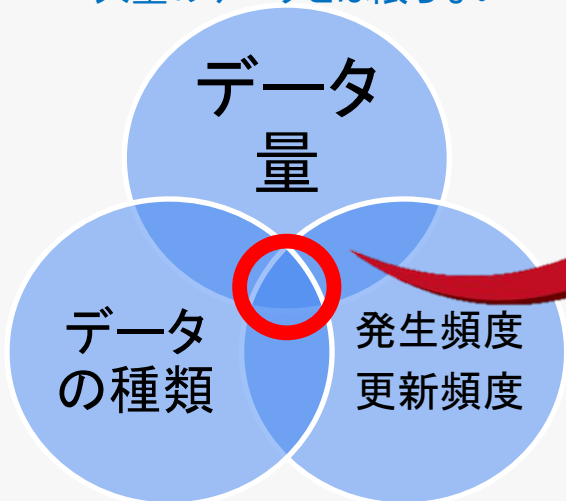
[次の領域へ](#)



ビッグデータとは？



ビッグデータとはその名の通り
大量のデータとは限らない



大量で多岐の種類のデータなのが

ビッグデータ

量	種類	速度
数TB～数PB	文書、画像、音声、動画	リアルタイム更新
		

マルチメディアとは

文字、音声、静止画、動画などの**複数のメディアが統合された**もの。

マルチメディアの利用目的は、「**コンテンツ**」(情報)を人から人へ伝達することである。

マルチメディア社会

マルチメディアの普及により、情報の内容と伝達の手段が増え、人と人との新たなコミュニケーションの手段が生まれた社会のこと

データベースとは



大量のデータを一つの場所に**集約し、保管・管理**し、さらに参照したいデータを簡単に抽出できるようにする**箱**のようなシステム。

データベースの管理



SQL

DBでデータの**参照・追加・削除**などを行うコマンド



整理

検索

更新

共有



DBMS

データベースの種類

リレーショナル型

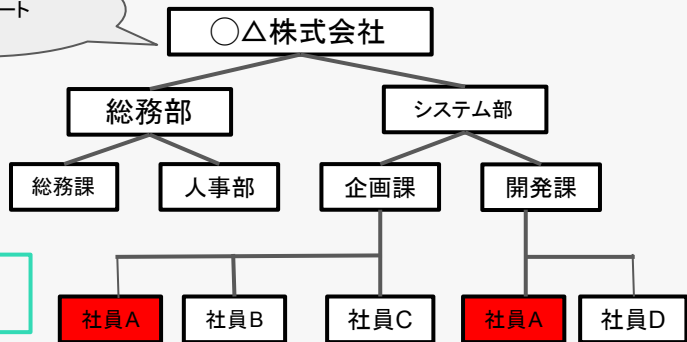
社員テーブル

社員番号	氏名	所属部署	兼務部署
001	社員A	0003	0004
002	社員B	0003	
003	社員C	0003	
004	社員D	0004	

部署テーブル

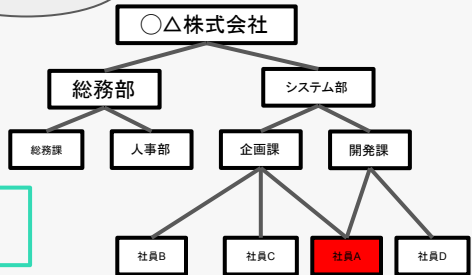
部署コード	部署名
0001	総務課
0002	人事部
0003	企画課
0004	開発課

検索のスタート



階層型

スタート位置は自由



ネットワーク型

データマイニングとは



大量のデータの中から色々な分析手法を駆使し、
有益な知識を採掘(マイニング)する技術。



知識発見(探索)

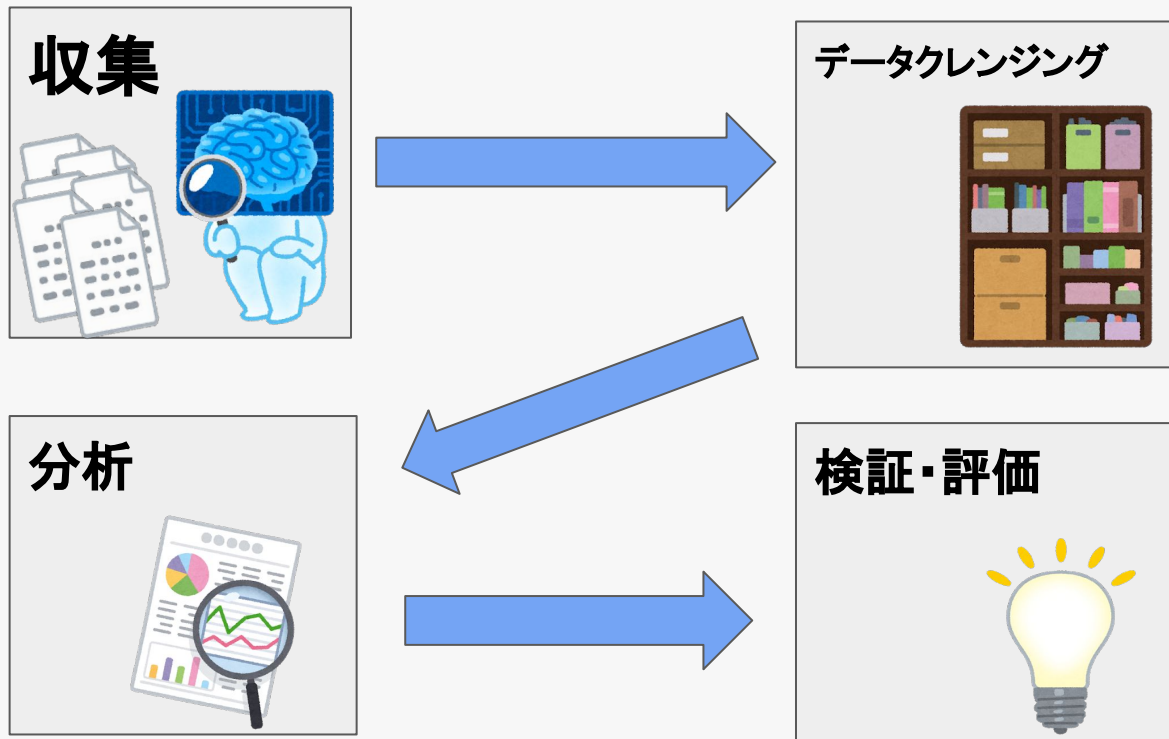
蓄積されたデータから、新しいパターン
やルールなどの知識を自動的に発見す
る。

ex) 機械学習、ディープラーニング

仮説検証

目的に沿って必要なデータを集め、分
析する。


データマイニングの手順



データウェアハウスとは



大量で多種多様な**データを時系列**で整理し保管しておく場所
やそれを管理するシステムのこと

1.  々が**時系列**である
- 2: **サブジェクトごと**に保管してある
- 3: データを**統合**してある
- 4: データは**消さない**

ロボティクスとは



ロボットの設計・製作・制御を行う「ロボット工学」を指す。

設計



機械工学

製作



電気電子工学

制御



情報工学

OLAPとは



オンライン分析処理のことである。

ここでいうオンラインとはオンライン授業などのオンラインという意味ではなく、ユーザーに素早く大量のデータから集計・分析した結果を届けることを指す。

大量のデータ



集計・分析



ユーザー



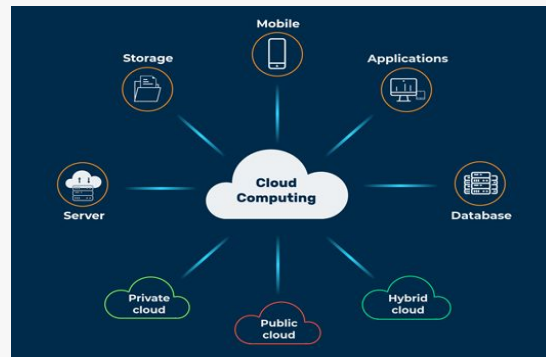
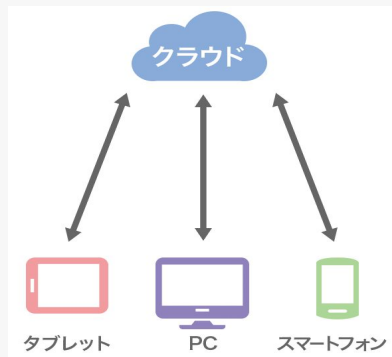
クラウド・コンピューティングとは？

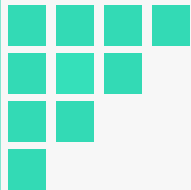


クラウドは雲を意味し、もともとネットワーク図でインターネットなどの外部のネットワークを雲の絵やアイコンで表現する。

クラウドコンピューティングとは、インターネット上のサーバーにあるコンピューターが提供している機能を、インターネット経由で利用する仕組みのこと。

わかりやすく言うと自身のPCにアプリケーションをインストールしなくてもインターネットに接続できる環境があれば、サーバー上で提供している機能やサービスを利用することができる。





人工知能学



[前の領域へ](#)

[マンダラートマップへ](#)

[次の領域へ](#)



チューリングテストとは

アラン・チューリングが提案した、ある機械が「**人間的**」かどうかを判定するためのテストである。

人間とコンピュータを分つものは何か、知的であるかどうか、人工知能であるかどうかを判断するために発明された

シンギュラリティとは



技術的特異点のことを言います。技術的特異点とは、人工知能（AI）が発達し、AIが**人間の知性を超え、人間の生活に大きな変化をもたらす**もの。

シンギュラリティが社会に与える影響とは、AIの技術が進歩することで自動化が普及し今まで人の手で行っていたものに人が関与しなくなることで **人的コストの削減**につながり、**デフレーション**を引き起こします。

イメージするなら...

映画のトランスフォーマー、スターウォーズなどに登場する AIロボットたち

機械学習とは



最も広く採用されている機械学習手法は、**教師あり学習**と**教師なし学習**の2つ。
機械学習の大部分(約70%)は、**教師あり学習**。
教師なし学習は、全体の10~20%を占める。そのほか **強化学習**という手法も使われている。

教師あり学習

教師あり学習とはあらかじめ用意された正解データを参照しながら、膨大なデータから特徴やルールを学習していく学習方法。

教師なし学習

教師なし学習とは正解データに頼らず、データそのものの特徴やルールを自ら見つけ出していく学習方法。

強化学習

強化学習とは、失敗や成功を繰り返し、それに対して報酬を与えることで学習効率を上げていく学習方法。

ディープラーニングとは

人間の神経細胞の仕組みを再現した **ニューラルネットワーク** を用いた **機械学習** の手法の1つであり、**多層構造のニューラルネットワーク** を用いることが特徴。

「**ニューラルネットワーク** (Neural Network:NN)」とは、人間の脳内にある神経細胞(ニューロン)とそのつながり、つまり神経回路網を人工ニューロンという数式的なモデルで表現したもの。

二つの違いに**注意**！

機械学習

学習する目的や内容については**人間が手を加える**

ディープラーニング (深層学習)

何を学ぶべきなのかも**機械が判断する**

ニューラルネットワーク

とは



脳の神経回路の一部を模した数理モデル

数理モデル...

現実の対象を簡略化し、物理法則にしたがって諸量の関係を数学的に表したもの。
とくに時間変化する現象を微分方程式などで記述することを指す。



自然言語処理とは



自然言語(話し言葉・書き言葉)を対象として言葉が持つ意味を様々な方法で解析・処理すること。

曖昧な言葉や意味が重複するものには弱い

処理行程



※イメージは完成済みのジグソーパズル



機械可読目録

コンピュータが語彙を理解する時に必要な辞書

コーパス

自然言語を収集し、構造情報を整理したデータベース

1.形態素解析

パズルをバラバラにして(文章を小さく区切り、情報を振り分ける)ピースごとの色や形を表す

3.意味解析

ピースの塊がパズルのどこにくるのか(文章を正しい解釈に分析する)確認する。

2.構文分析

ピースの凹凸を確認して(単語同士の関係性の解析・図式化)どれがくっつくか確認する。

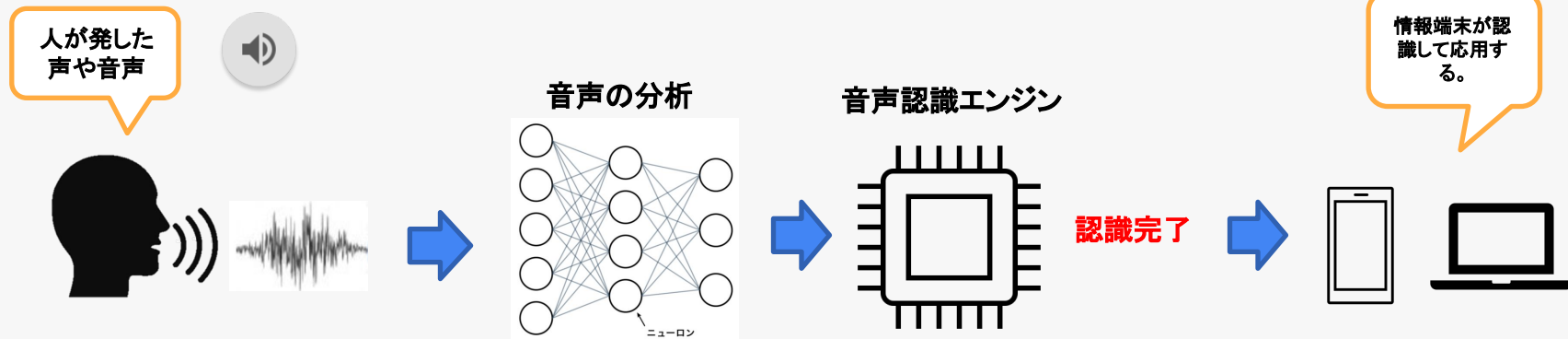
4.文脈解析

塊同士を組み合わせて完成させる。(複数の塊同士の関係性を分析する)

音声処理とは



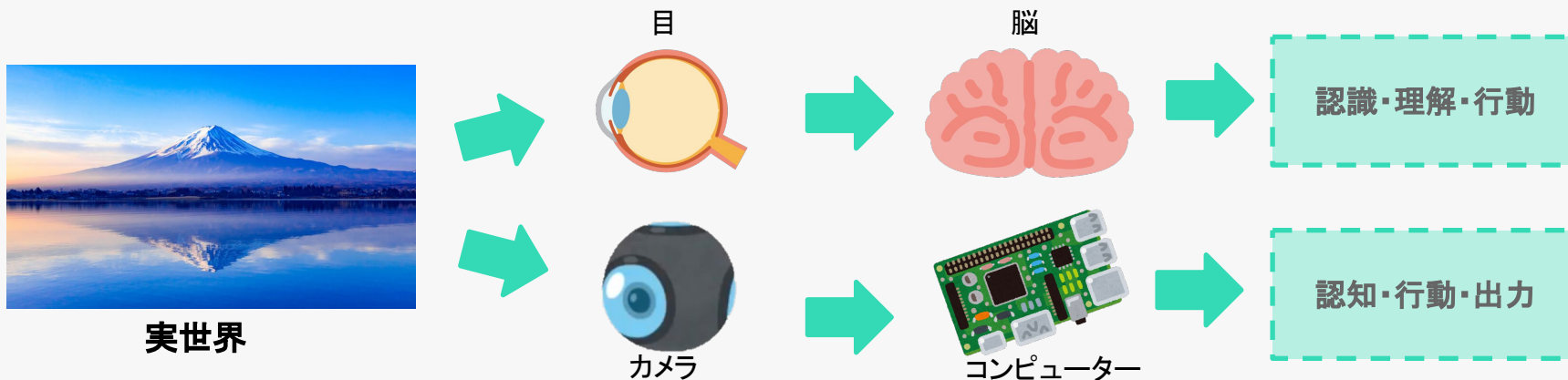
人が発した“音声”をコンピュータに“認識”させることを目指した技術領域



画像処理とは



コンピュータを利用して、**画像に情報を書き加えたりなど**の処理の全般を指す



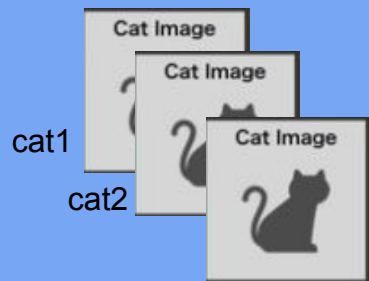
最新の人工知能との組み合わせによりコンピューターに様々な能力を実装することができる。

人間を超える画像処理技術も多く存在 → **できることはたくさんある**

画像処理をする上で画像を認識する必要がある

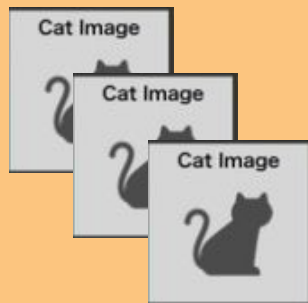


教育あり学習



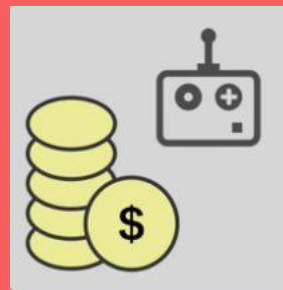
入力データと正解ラベルをセットで学ばせる学習法

教育なし学習

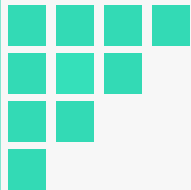


入力データだけを学ばせる学習法

強化学習



コンピューターが自ら学び、報酬を得る学習法



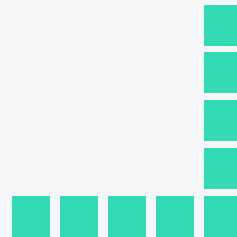
情報学



[前の領域へ](#)

[マンダラートマップへ](#)

[次の領域へ](#)



情報技術(IT)とは

IT

情報を**取得、加工、保存、伝送**するための科学技術のことである。ITは医療、交通、農業、自動車の分野などで活用されている。

ICT

情報技術だけではなく、その情報や技術を共有するための「**コミュニケーション**」の意味がより強調された用語である。

IoT

身の回りの**あらゆるモノがインターネットにつながる仕組み**を表す。

セキュリティとは



データやシステム、**通信経路の保護**や外部からの攻撃・改竄などの**危険を排除**すること

コンピュータセキュリティ



災害や不正アクセスなどからコンピュータを守り、**ハードウェア、ソフトウェア、データ、ネットワークのCIA**を維持を目的とする。

セキュリティ対策



物理的対策

耐震、防火、UPS、バックアップ

論理的対策

アクセス認証、ウィルス対策、不正アクセス 防止策など

ソフトウェアとは

コンピューターを動かすプログラムでコンピューター自体や周辺機器などのように目に見える機器をハードウェアと呼ぶのに対し、**ソフトウェアは目で見ることができない。**

ソフトウェアの種類は、**基本ソフトウェアのOS(オペレーティング・システム)と応用ソフトウェアのアプリケーションソフトウェア**の2つに大きく分類される。



OS(オペレーティングシステム)

OS(オペレーティング・システム)は、**アプリケーションが動作するベース**となるソフトウェアです。基本ソフトウェアとも呼ばれる。

OSの役割

OSの大きな役割は、**アプリケーションソフトウェアが動く環境を作ること**。

その他にもコンピュータのハードウェアの管理、利用率の向上、ユーザーインターフェースの提供、ファイルやタスクの管理など幅広い役割を担う。



アプリケーションソフトウェア

アプリケーションソフトウェアとは、**コンピュータ上で使いたい機能を実現する**ソフトウェアでアプリやソフトとも呼ばれる。

簡単に言うと**OS上で具体的にやりたいことを実現するためのソフトウェア**。

アプリはパソコンにインストールするものと、Webブラウザ上で動作するWebアプリを2つ分けられる。

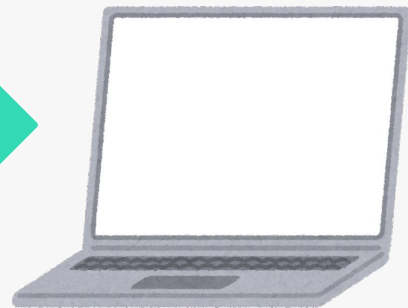
OSに合わせてアプリケーションソフトウェアは作られるため、違うOS上では動作できない。



サーバーとは



ネットワークを介して、他のコンピュータに
サービスや情報を提供するコンピュータのこと



情報

アルゴリズムとは



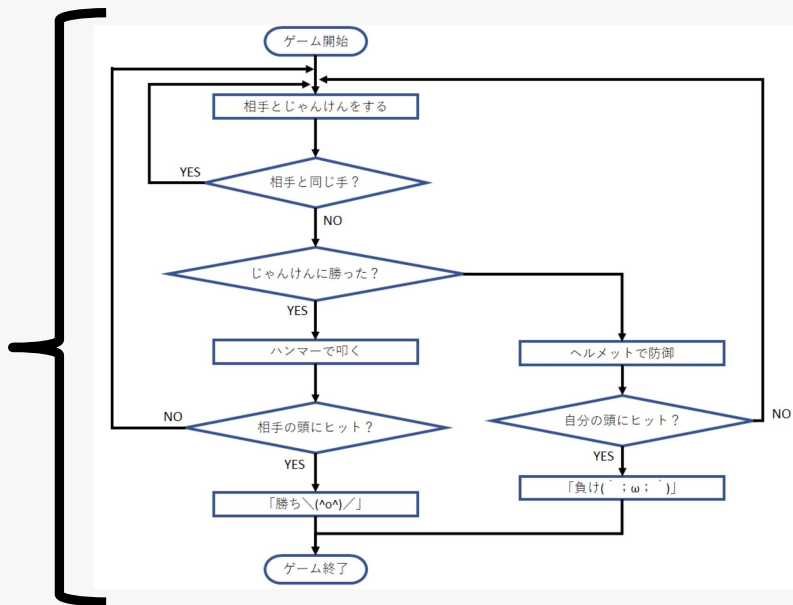
特定の問題を解く手順を、**単純な計算や操作の組み合わせとして明確に定義**したもの

データサイエンス

こんなことがしたい
＜目的＞

アルゴリズム
組み立て

データ収集
データ解析



アルゴリズムを
情報的に表現

フローチャート

処理のアルゴリズムを視覚的に表現できる図解法です。プログラムの設計によく用いられます。

フィンテックとは

Finance(金融) × Technology(技術)



金融サービスとテクノロジーを結びつけ、会計業務の効率化や顧客体験の価値向上を図る取り組み

ヒューマンエラーとは



人間が原因で起こる失敗、人為的ミス

意図したヒューマンエラー

本人はまったく意図していないのにミスをしてしまうタイプのヒューマンエラー

手順無視

疲労

手抜き

集中力の低下



意図していないヒューマンエラー

本人がある程度の意図を持って行動したが、起きてしまうヒューマンエラー

知識不足

連絡連携不足・勘違い

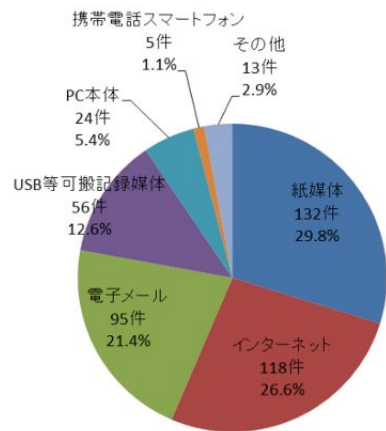
不慣れ



ヒューマンエラーの情情的問題



情報学におけるヒューマンエラーは**情報漏洩**があげられる



デー

タ:<https://tokiocyberport.tokiomarine-nichid.o.co.jp/cybersecurity/s/column-detail/17>

2017年
(N=386件)

2018年
(N=443件)

紙媒体
(150件)

紙媒体
(132件)

インターネット
(87件)

インターネット
(118件)

電子メール
(77件)

電子メール
(95件)

USB等可搬
記録媒体
(41件)

USB等可搬
記録媒体
(56件)

紙媒体による漏えい件数が
最も多い

インターネット・電子メール・USB
経由が増加

また、近年ではコンピューターウイルスやフィッシング詐欺などの特殊な方法でじょうほうが漏洩しているパターンもある。

環境の更新



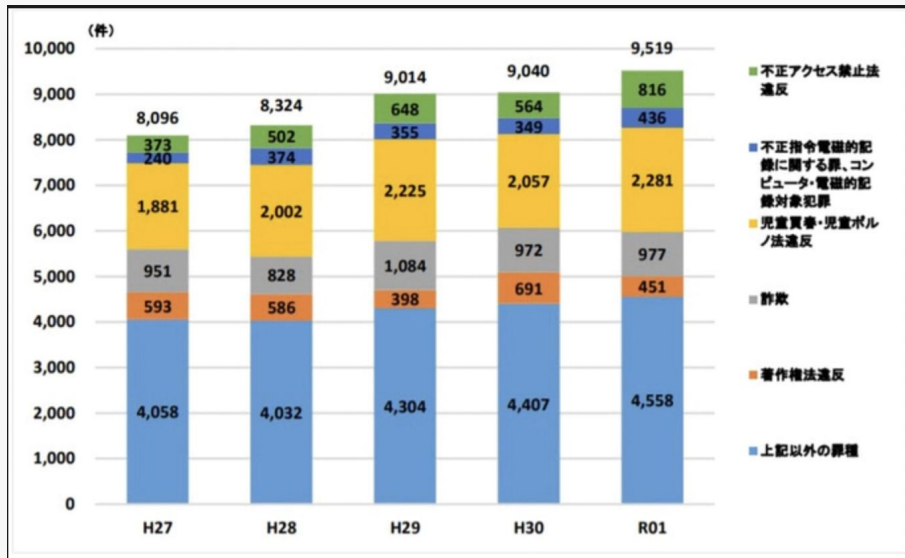
<https://thebest-1.com/a5147/>

ウイルス対策ソフトの導入

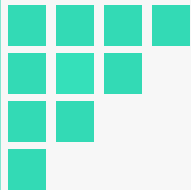
サイバー犯罪とは



コンピューター、コンピューターネットワーク、またはネットワークに接続された端末を標的とした犯罪行為



サイバー犯罪の件数



プログラミング



[前の領域へ](#)

[マンダラートマップへ](#)

[次の領域へ](#)



関数

数学における関数

二つの変数 x 、 y があって、 x の値が決まれば、それに伴って y の値がただ一つ決まるとき、 y は x の**関数**であるという。

プログラミングにおける関数

関数とは、コンピュータプログラム上で定義されるもので数学の関数のように与えられた値(**引数**)を元に何らかの計算や処理を行い、結果を呼び出し元に返すもののこと。

ソースコードとは



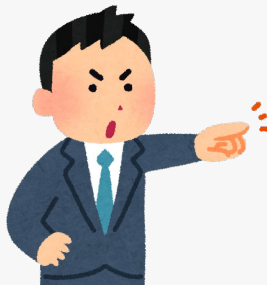
ソースコードとは、プログラムを動作させるため、**プログラミング言語を使って記述された指示書**のこと。

コンピュータはソースコードに書かれた処理手順いわゆる「命令」に従い、上から下へと順に実行する。

プログラミング言語に限らず、人工言語や一定の規約・形式に基いて記述された複雑なデータ構造の定義・宣言などのことも**ソースコードと呼ぶ**。



ソースコードを書く



コンピュータに命令する



コンピュータがソースコードを実行する

オブジェクトコード



コンピュータは2進数の数値のデータ、いわゆる**機械語しか読めない**。

ソースコードのプログラムはそのまま実行できない→機械語に変換しないと！

変換後の機械語によるプログラムを

「オブジェクトコード」(object code)

「オブジェクトプログラム」(object program)

「ネイティブコード」(native code)

「ネイティブプログラム」(native program)

「バイナリコード」(binary code)

などと呼ぶ。



オープンソースとは



商用、非商用の目的に問わず利用、修正が可能なコード・ソフトウェア

概念

オープンソースの概念をソースに例えると、

「ソース（商品）のレシピが公開されている」ということ。

一般的なA社では、

レシピは **秘密**



一般的なB社では、

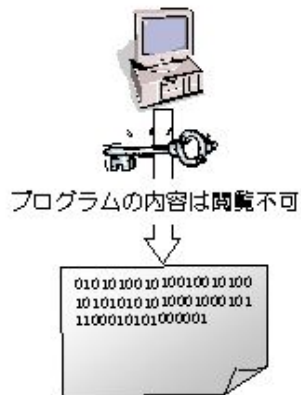
レシピを **公開**



オープンソース活用までの流れ

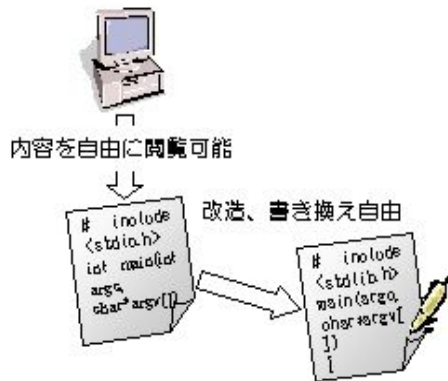


一般市販のコンピュータプログラム



- ・内容が分からないので、改造、書き換え不可能
- ・改良はメーカー任せ

オープンソース



- ・中身を触って、使いやすく改良、書き換えが可能
- ・自分自身で対処も可能

代表的なもの



オブジェクトとは

オブジェクトとはコンピュータ上で操作や処理の対象となる**何らかの実体**のこと

機械語に変換されたソースコードのことを**オブジェクトコード**という

ここでいうオブジェクトとはモノではなく目的という意味である。

オブジェクトコードを格納したファイルを**オブジェクトファイル**という。

オブジェクト指向ー1

オブジェクト指向とは
様々なオブジェクトを組み合わせて関連性や相互作用を記述していくこと
により、システム全体を構築していく手法のこと

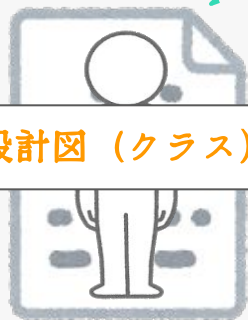
そして、関連するデータと処理手順をオブジェクトと呼ばれる
一つのまとまりとして定義する。

オブジェクト指向ー2

人間の設計図は...

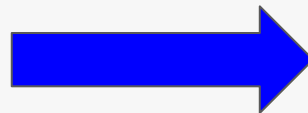


設計図 (クラス)



プロパティ(属性)
オブジェクトの状態を示す

脳の大きさ
指の数
目の位置 etc...



オブジェクト



メソッド(操作)
オブジェクトがアクションを起こすときの一つ
一つの処理

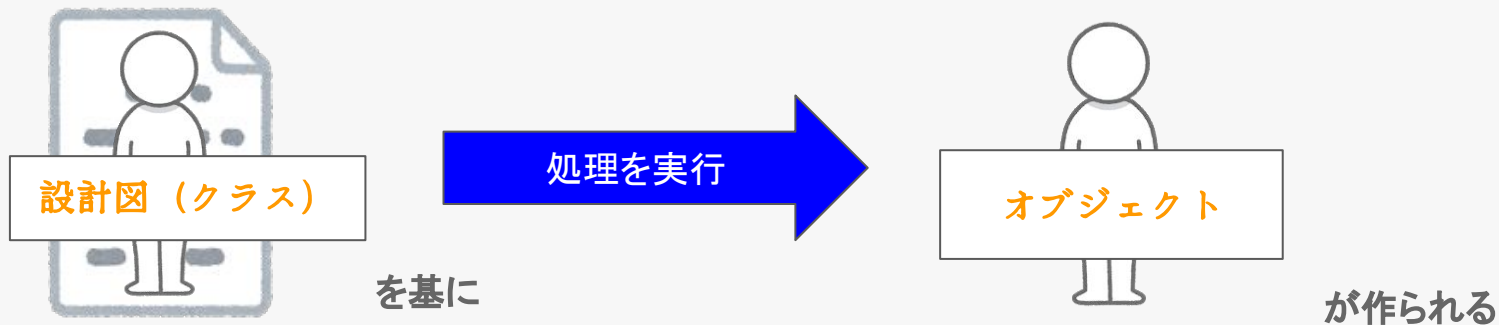
脳の働き、関節の曲がり具合、目の動き

メソッドとは



オブジェクト指向プログラミングにおいて

メソッドとは
設計図のパーツ(機能)、一つ一つの設計図



メソッドはExcel(イメージ)



sumというメソッド(機能)

sum(2, 4)



sumくん
2+4してね

Excelの表

(2, 4)

はい、データ(2
と4)を受け取り
ました

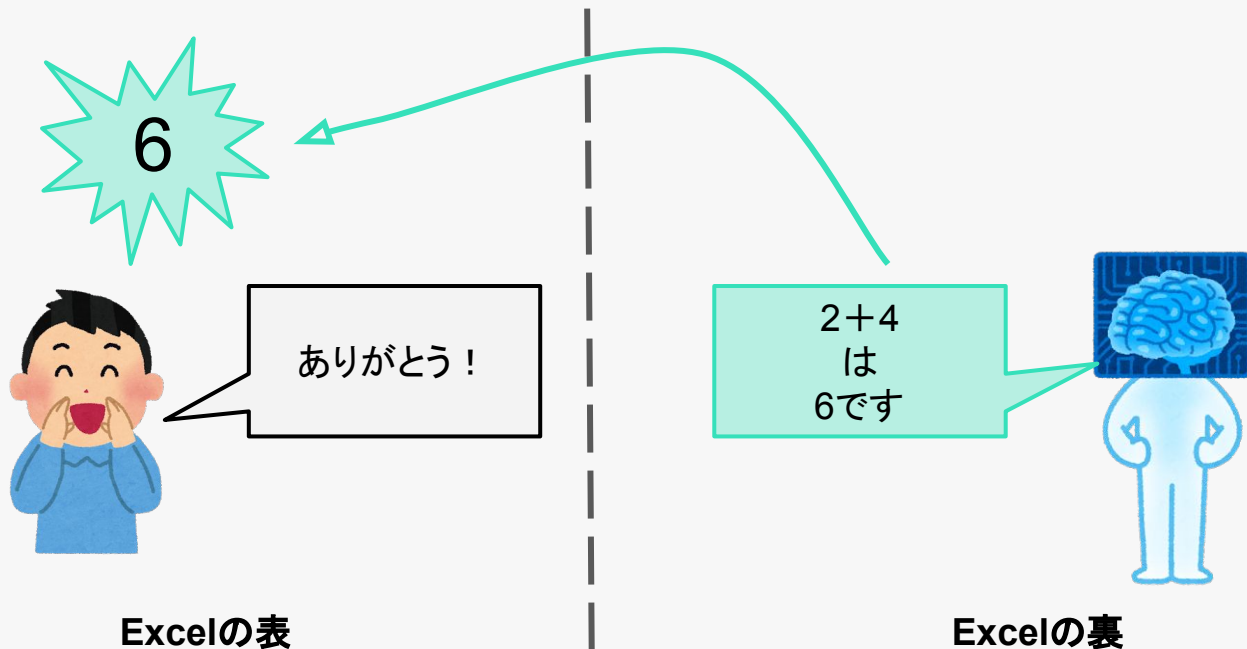


Excelの裏

メソッドはExcel(イメージ)



処理は隠して結果を表示



エラーとは

とは、コンピュータにおいて、プログラムがそれらの実行が正常な動作でないと判断し、処理を中断または停止させる状態。

代表的なエラー

- 1.Operation System not found
- 2.ntldr is missing
- 3.disk boot failure



数学



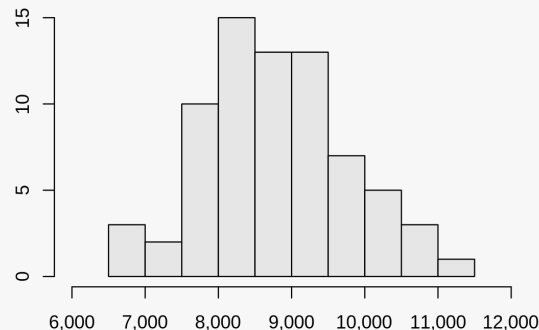
[前の領域へ](#)

[マンダラートマップへ](#)

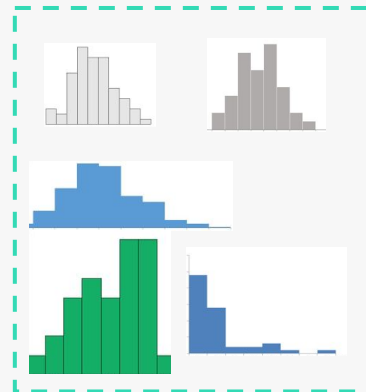
[出典一覧へ](#)

ヒストグラムとは

- ・縦軸を「度数」、横軸を「階級」
として、データを長方形で表したものの。
- ・データがとても多いもの
→階級に分けているから見やすい。
→どの階級の度数が高いかが一目で分かる。



こんなデータがたくさんあったら
どうやって比較するの？
→「代表値」を使って比較する！

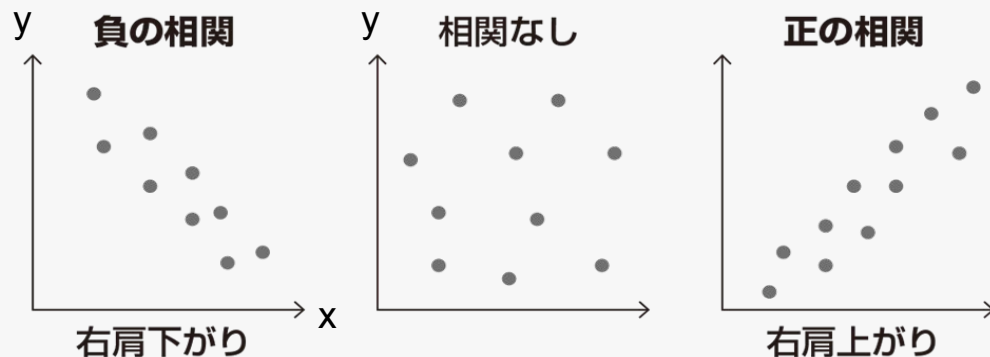


相関関係とは



相関関係とは、AとBの事柄になんらかの**関連性**があるものを表すもの。

相関関係には「**正の相関**」と「**負の相関**」がある。



図のように、**正の相関関係**とは横軸の値(x)が増加すると縦軸の値(y)も増加するという関係のに対し、**負の相関**とは x が増加すると y が減少するという関係。

x と y は全然関係がない場合もある。

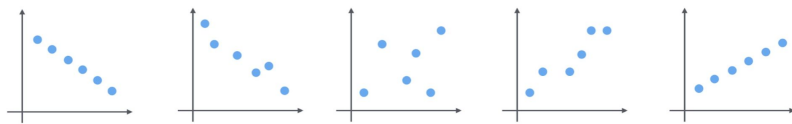
相関係数



相関係数は2種類のデータとの**相関関係**を表す値。

ExcelやGoogle Spreadsheetsで、**Correl関数**を使うことで**相関係数**が求められる。

$$\text{Correl}(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$



-1 -0.5 0 0.5 1

強い負の
相関関係

相関関係なし

強い正の
相関関係

	A	B	C
1	Data1	Data2	
2	3	9	
3	2	7	
4	4	12	
5	5	15	
6	6	17	
7			
8	数式:	=CORREL(A2:A6,B2:B6)	
9	結果:	0.997054486	

相関係数は -1 から 1 までの値を取る。

相関係数は -1 に近いほど、強い負の関係とされる。逆に 1に近いほど、強い正の関係とされる。

分散とは

データの散らばりの度合いを表す値のこと。

データの値が x_1, x_2, \dots, x_n で、この平均値が \bar{x} のとき、次の値を分散と呼ぶ

$$\frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}$$

平均値

データの個数

期待値とは



1回の試行で得られる値の平均値のこと。また、期待できる値のこと

ある台に1,000円を投入したら平均して
1,200円が返ってくる台があったとする。



期待値は200円となる

1,000円を投入したら平均して800円
が返ってくる台があったとする。



期待値は-200円となる

期待値の計算方法



例題1:

さいころを1回投げるとき、出る目の期待値はいくらになるでしょうか。



出る目(x)	1	2	3	4	5	6
確率(p)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$$

$$= 3.5$$

さいころを何回か投げると1回当たりの出る目の平均は3.5になると期待できます

例題2:

さいころを投げて、出た目の数×100円貰えるゲームをします。このゲームを1回やるとき、いくら貰えると考えられるでしょうか。



出る目	1	2	3	4	5	6
貰える金額(x)	100円	200円	300円	400円	500円	600円
確率(p)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$= 100 \times \frac{1}{6} + 200 \times \frac{1}{6} + 300 \times \frac{1}{6} + 400 \times \frac{1}{6} + 500 \times \frac{1}{6} + 600 \times \frac{1}{6}$$

$$= 350$$

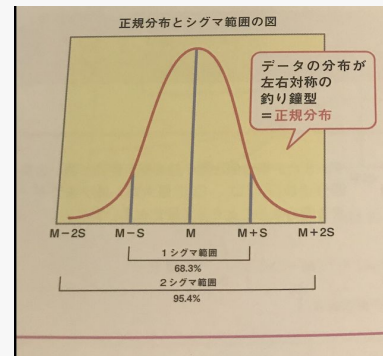
この結果から、「このゲームを1回行うごとに350円くらいもらえると見込んでよい(期待できる)」

標準偏差とは

分散は測定値を2乗するため、単位も元の単位とは異なり解釈が難しい。その点を補ったものが標準偏差である。

$$\sqrt{\frac{1}{n} \left\{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right\}}$$

標準偏差の読み方の目安として次のようなものがある。データ分布が左右対称で釣鐘型をする場合（より正確に言えば正規分布をなす場合）M-S以上、M+S以下の範囲に全データの約68.3%。M-2S以上、M+2S以下の範囲に全データの約95.4%。M-3S以上、M±3S以下の範囲に全データの約99.7%が含まれている。



平均値とは

平均または平均値とは、数学において、数の集合やデータの中間的な値のこと。

メリット

平均値は、すべてのデータを足して、データの個数で割るという過程を経ているので、「すべてのデータを考慮できている」というのがメリット。

デメリット

平均値のデメリットとしては、「異常値(極端なデータ・値)に左右されやすい」ということ。

(5, 6, 7) の平均は？

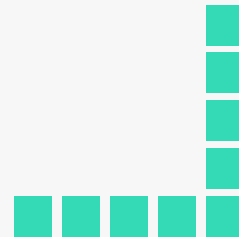
$$\text{平均} = \frac{\text{合計 } 5+6+7}{\text{個数 } 3} = 6$$



出典一覧

[前の領域へ](#)

[マンダラートマップへ](#)



推計統計学

出典: 大学4年間の統計が10時間で学べる 東京大学教養学部教授倉田博史 P38,P39

ISBN978-4-04-604187-6

疫学統計

Jミルク ちょっと気になる疫学研究って? (参照日 2021/05/16)

https://www.j-milk.jp/knowledge/food-safety/uwasa_ekigaku.html

science shift 疫学入門～疾病の原因と病の関係性を明らかにするための考え方 (参照日2021/05/16)

<https://scienceshift.jp/epidemiology/#heading1>

多変量解析

スライドに記載

ベイズ統計学

スライドに記載

要約統計学 スライドに記載

記述統計学

記述統計学と推計統計学の違い <<https://ai-trend.jp/basic-study/basic/inductive-statistics/>> (参照: 2021/5/9)

馬場真哉、記述統計の基礎 <<https://logics-of-blue.com/記述統計の基礎/>> (参照: 2021/5/16)

記述統計学と推測統計学の違い、

<<https://toukeigaku-jouhou.info/2018/02/27/descriptive-statistics-and-inferential-statistics/>> (参照: 2021/5/16)

社会統計学

What is Social Statistics ? <<https://www.socialsciences.manchester.ac.uk/social-statistics/study/what-is-social-statistics/>> (参照: 2021/5/9)

Social statistics definition, <<https://sociologydictionary.org/social-statistics/>>, Open Education Sociology Dictionary, (参照: 2021/5/9)

統計応用

出典: <https://core.ac.uk/download/pdf/293133421.pdf>

ビッグデータ

<https://data.wingarc.com/what-is-big-data-11866>

<https://ledge.ai/big-data/>

データベース

<https://proengineer.internous.co.jp/content/columnfeature/6411>

<https://www.idcf.jp/words/database.html>

<https://news.mynavi.jp/kikaku/20210322-1813647/>

<https://it-trend.jp/database/article/89-0070>

データマイニング

<https://ledge.ai/data-mining/>

<https://youtu.be/AzGcySYRT34>

データウェアハウス

<https://it-trend.jp/dwh/article/149-0001#chapter-4>

<https://www.cross-m.co.jp/column/marketing/mkc20210305/>

<https://boxil.jp/mag/a1657/>

ロボティクス

<https://www.keyence.co.jp/ss/general/iot-glossary/robotics.jsp>

<https://eiga.com/news/20201029/19/>

OLAP

出典: <https://it-trend.jp/dwh/article/149-0010>

クラウドコンピューティング

[クラウド・コンピューティングとは？ | 用語集 | SBクラウド株式会社](#) 閲覧日: 2021/5/22

[クラウドコンピューティングってどんなもの？ 定義やメリットをわかりやすく紹介\(初心者向け\)](#) 閲覧日: 2021/5/22

機械学習

【保存版】機械学習とは | 意味や仕組み・勉強方法を徹底解説、<<https://ledge.ai/machine-learning/>> (参照: 2021/5/28)

機械学習、<https://www.sas.com/ja_jp/insights/analytics/machine-learning.html> (参照: 2021/5/28)

機械学習とは、<<https://www.data-artist.com/contents/machine-learning.html>> (参照: 2021/5/28)

機械学習と深層学習の違いとは？メリットや課題を挙げながら解説、<<https://www.dsk-cloud.com/blog/difference-machine-learning-and-deep-learning>> (参照: 2021/5/28)

「教師あり学習」と「教師なし学習」は何が違う？イラストでわかる機械学習の基礎、<<https://www.sbbit.jp/article/cont1/49067>> (参照: 2021/5/28)

ディープラーニング

機械学習と深層学習の違いとは？メリットや課題を挙げながら解説、
<<https://www.dsk-cloud.com/blog/difference-machine-learning-and-deep-learning>> (参照: 2021/5/28)

深層学習とは、<<https://www.ntt.com/bizon/glossary/j-s/deep-learning.html>>、(参照: 2021/5/28)

ニューラルネットワーク

<https://ledge.ai/neural-network/>

出典: 武蔵野大学ムサシノスカラシップ選抜エントリーシート

チューリングテスト

<https://ledge.ai/turing-test/#:~:text=%E3%82%82%E3%81%97%E4%BA%BA%E9%96%93%E3%81%AE%E3%82%88%E3%81%86%E3%81%AA.%E4%BD%BF%E3%82%8F%E3%82%8C%E3%81%A6%E3%81%8D%E3%81%BE%E3%81%97%E3%81%9F%E3%80%82>

自然言語処理

https://ainow.ai/natural_language_processing/

<https://ledge.ai/nlp/>

音声処理

<https://iphone-mania.jp/news-205564/>

<https://www.youtube.com/watch?v=OX2-tpTkLA0>

<https://ledge.ai/speech-recognition/>

https://www.bcm.co.jp/solution-now/cat-solution-now/2018-12_1522/

<https://thinkit.co.jp/story/2015/09/16/6419>

<https://thinkit.co.jp/story/2015/09/16/6419>

画像処理

https://www.youtube.com/channel/UCbCHn_q5MUAz5ZDNA-wKsug?app

<https://ledge.ai/image-recognition/>

<https://ainow.ai/2019/07/11/173264/>

<https://time-space.kddi.com/mobile/20181023/2473>

<https://base.exawizards.com/view/modelDetail?id=16>

<https://nomad-ceo.com/ios-12>

シンギュラリティ

https://job.mynavi.jp/career_tanq/articles/?id=145

<https://bizhint.jp/keyword/42911>

情報技術

<https://ja.wikipedia.org/wiki/%E6%83%85%E5%A0%B1%E6%8A%80%E8%A1%93>

セキュリティ

<https://business.ntt-east.co.jp/content/cloudsolution/column-154.html>

<https://www.idcf.jp/words/security.html>

ソフトウェア

パソコンのソフトウェアとは何なのか分かりやすく解説・種類や役割について、
<<https://tech-camp.in/note/technology/86909/>> (参照: 2021/5/31)

ソフトウェアって何?、<<https://employment.en-japan.com/tenshoku-daijiten/14870/>> (参照: 2021/5/31)

アルゴリズム

<https://digitalidentity.co.jp/blog/seo/seo-tech/ranking-and-session.html>

<https://www.python.jp/pages/about.html>

http://research.nii.ac.jp/~uno/algo_3.htm

<https://codezine.jp/article/detail/9900?p=2>

<https://basics.k-labo.work/2017/10/02/%E3%82%A2%E3%83%AB%E3%82%B4%E3%83%AA%E3%82%BA%E3%83%A0%E3%81%A8%E6%B5%81%E3%82%8C%E5%9B%B3/>

https://recruit.gmo.jp/engineer/jisedai/blog/airl_with_open_ai_gym/

ヒューマンエラー

<https://tokiocyberport.tokiomarine-nichido.co.jp/cybersecurity/s/column-detail17>

<https://thebest-1.com/a5147/>

サイバー犯罪

<https://www.colorfulbox.jp/media/server/>

関数

コトバンク 関数とは（参照日2021/06/17）

<https://kotobank.jp/word/%E9%96%A2%E6%95%B0-2523>

IT用語辞典 関数(プログラミング) (参照日 2021/06/17)

<https://e-words.jp/w/%E9%96%A2%E6%95%B0.html>

オブジェクト

オブジェクト指向とは？

<https://eng-entrance.com/what-oop> (参照2021-06-16)

オブジェクト(object)とは

<https://e-words.jp/w/オブジェクト.html> (参照2021-06-16)

エラー

https://www.seikatsu110.jp/pc/pc_computer/31355/

ソースコード

ソースコードとは？～初心者のための基本解説～、< <https://www.homepage-seisaku.jp/post61/#heading1> >
(参照2021/6/17)

ソースコード【source code】ソースプログラム / source program / ソースリスト < <https://e-words.jp/w/ソースコード.html> > (参照2021/6/16)

オープンソース

<http://research.nii.ac.jp/~uno/struct.htm>

<https://www.orca.med.or.jp/orca/summary/opensource.html>

<https://www.ossnews.jp/oss/>

メソッド

<https://e-words.jp/w/メソッド.html> (参照2021/6/19)

https://youtu.be/W_IHfDqt5EI (参照2021/6/19)

ヒストグラム

大学4年間の統計学が10時間で学べる p10~p11

東京大学教養学部教授 倉田博史 ISBN 978-4-04-604187-6

分散

【高校 数学 I】データの分析 10 分散とは

<https://youtu.be/dOnJpfvTAKs> (参照2021-06-16)

【基本】データの分散 | なかけんの数学ノート

<https://math.nakaken88.com/textbook/basic-variance-in-data/> (参照2021-06-16)

標準偏差

大学4年間の統計学が10時間で学べる p10~p11

東京大学教養学部教授 倉田博史 ISBN 978-4-04-604187-6

相関関係

「相関関係」「因果関係」の違いとは？見分けるポイントと事例をわかりやすく解説！

<<https://ferret-plus.com/6461>> (参照: 2021/6/17)

26-2. 正の相関と負の相関 <<https://bellcurve.jp/statistics/course/9589.html>> (参照2021/6/17)

CORREL 関数 <<https://support.microsoft.com/ja-jp/office/correl-関数-995dcef7-0c0a-4bed-a3fb-239d7b68ca92>> (参照: 2021/6/17)

5分でわかる！「正の相関」「負の相関」と「相関係数」

<<https://www.try-it.jp/chapters-6303/sections-6304/lessons-6353/>> (参照: 2021/6/17)

トレンドライン: 線形回帰の引き方

<https://exploratory.io/note/GMq1Qom5tS/14-2-rdp7QBM4vw/note_content/note.html> (参照: 2021/6/17)