

aa-3. 実データによるデータサイエンス・AI の演習

(人工知能)

金子邦彦



3-1 オープンデータ

オープンデータ



誰もが自由に利用できるデータ

オープンデータの定義

誰でも自由にアクセスし、
利用、加工、再配布できるように
公開されたデータ

アクセス可能
(Accessible)

自由な利用・加工

自由な再配布

提供元：



政府



地方自治体



企業



研究機関



個人

普通のデータとの違い

	普通のデータ	オープンデータ
1. 形式	人向け、独自形式 (e.g., PDF, 紙)	コンピュータ処理が容易なデジタル データ (e.g., CSV, JSON, XML)
2. ライセンス	著作権や二次利用が 不明確・制限	著作権やライセンスの制約が明示 され、二次利用が許諾されている
3. コスト	有償、取得に費用が かかる	原則として無償で取得可能
4. 連携	独自形式のため 連携が困難	標準化された形式により 他データとの連携・統合が容易

生活や社会に関する身近な例

気象庁の
気象・地震データ



天気予報アプリ、
防災情報の基盤

国土地理院の
地図データ



カーナビ、地図
アプリに利用

政府統計e-Stat



人口統計、経済指標

GTFS形式の
公共交通データ



乗換案内アプリ

自治体のハザードマップ、
AED設置場所、
ごみ収集日データ



地域の安全・生活情報

厚生労働省の
感染症動向データ



感染症の流行状況
把握、対策

日本の公的オープンデータ①

1. 気象庁

気象・地震データ (気象庁)



A



気象庁

気象庁公式ポータル

B

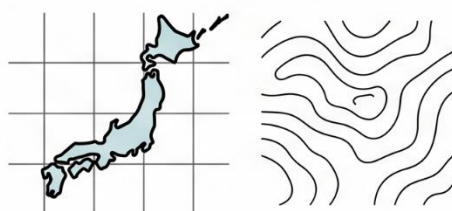


地震・津波データ

震源情報・津波観測記録

2. 国土地理院

地図データ (国土地理院)



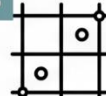
A



地理院地図

地形図・写真・標高・災害情報

B

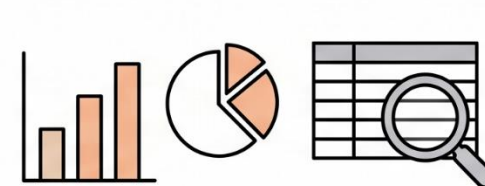


基盤地図情報

行政界・道路縁・建築物
外周線・標高点

3. 政府統計 e-Stat

政府統計 (e-Stat)



A



e-Stat

政府統計の総合窓口

B



統計DB検索

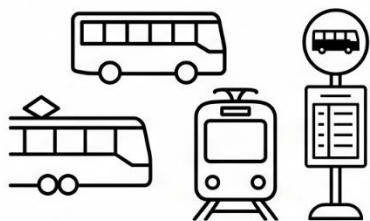
統計表・データベース
検索機能

気象庁は気象観測値および地震・津波の観測解析データを公開している。
国土地理院は電子国土基本図および基盤地図情報（基本項目・数値標高モデル等）を提供している。
政府統計e-Statは各府省の統計データを一元的に検索・取得できる総合窓口である。

日本の公的オープンデータ②



4. 公共交通データ (GTFS)



A 国交省 GTFS-JP案内

鉄道・バス等の運行情報
標準データ仕様

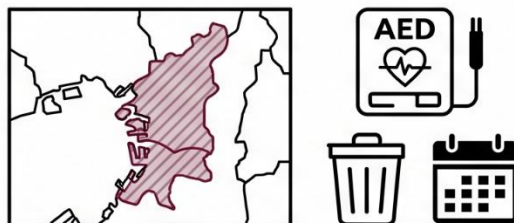


B GTFSデータリポジトリ

全国事業者のGTFSデータ
登録・公開基盤



5. 自治体オープンデータ



A 大阪市 水害ハザードマップ

洪水・内水・高潮・津波の
浸水想定区域

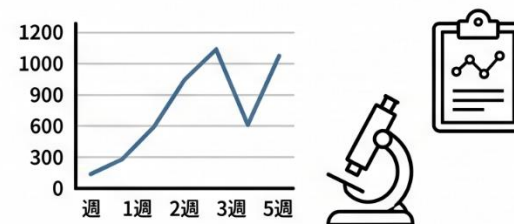


B 世田谷区 自府庁 自治体標準オープン データセット

AED設置箇所・ごみ分別・収集日等



6. 感染症動向 (厚生労働省)



A 感染症発生動向調査

感染症法に基づく医
療機関からの届出集計



B 感染症情報 (総合)

結核・その他感染症に
関する統計と資料



公共交通分野では、国土交通省がGTFS-JP仕様を策定し、鉄道・バス・旅客船等の運行情報を標準形式で提供する枠組みを整備している。GTFSデータリポジトリは事業者が同形式のデータを登録・公開する共通基盤である。

自治体分野では、大阪市が洪水・内水氾濫・高潮・津波に対応する水害ハザードマップを公開し、世田谷区は自治体標準オープンデータセットに従いAED設置箇所やごみ収集に関するデータを公開している。

保健分野では、厚生労働省が感染症発生動向調査を通じて感染症法に規定された疾患の発生状況を集計し、結核を含む各種感染症の統計情報を提供している。

政府統計 e-Stat



e-Stat は**政府統計の総合窓口**であり、**日本の統計**が閲覧できる政府統計ポータルサイトである

URL : <https://www.e-stat.go.jp/>

The screenshot shows the e-Stat website interface. At the top left is the 'e-Stat' logo with the tagline '政府統計の総合窓口'. To its right is the text '統計で見る日本' and 'e-Statは、日本の統計が閲覧できる政府統計ポータルサイトです'. On the right side of the header, there are links for 'お問い合わせ', 'ヘルプ', and 'English', along with a 'ログイン' button. Below the header is a navigation bar with links: '統計データを探す', '統計データの活用', '統計データの高度利用', '統計関連情報', and 'リンク集'. The main content area is divided into two columns. The left column features a '統計データを探す' section with a sub-header '(政府統計の調査結果を探します)' and a 'その他の絞込' button. It contains three large buttons: 'すべて' (with a bar chart icon), '分野' (with a cube icon), and '組織' (with a building icon). Below these is a search bar with the text 'キーワード検索 : AED' and a '検索' button. The right column features a '統計データの高度利用' section with a '利用ガイド' button and a 'マイクロデータの利用' button. Below that is a '開発者向け' button. At the bottom of the right column is a '統計関連情報' section with a '統計分類・調査計画等' button. The bottom of the left column features a '統計データを活用する' section with four buttons: 'グラフ' (with a bar chart icon), '時系列表' (with a clock icon), '地図' (with a map icon), and '地域' (with a location pin icon).

e-Stat
政府統計の総合窓口

統計で見る日本
e-Statは、日本の統計が閲覧できる政府統計ポータルサイトです

お問い合わせ | ヘルプ | English
ログイン

統計データを探す | 統計データの活用 | 統計データの高度利用 | 統計関連情報 | リンク集

● **統計データを探す** (政府統計の調査結果を探します) その他の絞込

すべて
政府統計一覧の中から探します

分野
17の統計分野から探します

組織
統計を作成した府省等から探します

キーワード検索 : AED 検索

● **統計データを活用する**

グラフ
主要指標をグラフで表示
(統計ダッシュボード)

時系列表
主要指標を時系列表で表示
(統計ダッシュボード)

地図
地図上に統計データを表示
(統計GIS)

地域
都道府県、市区町村の
主要データを表示

利用ガイド

● **統計データの高度利用**

マイクロデータの利用
公的統計のマイクロデータの利用案内

開発者向け
API、LODで統計データを取得

● **統計関連情報**

統計分類・調査計画等

政府統計データの概要

- 公表分野：人口，労働，経済，環境など
- 用途：ビジネス戦略，政策立案，研究，学習に役立つデータである

総務省統計局の統計表

- 公開されている統計表
 - 人口
 - 労働力
 - 国民経済
 - 物価
 - 社会生活基本調査 など多数
- ダウンロードページ：

<https://www.stat.go.jp/data/guide/download/index.html> 7

日本政府の家計調査



全国約9千世帯を対象として，家計の収入・支出，貯蓄・負債などを毎月調査するもの

「プリン」にお金を使うことが多い都市は？

	A	B	C	D	E	F	G	H	I	J	K
10											
11									/品目分類	1.1E+07	
12	表章項目	表章項目	世帯区分	世帯区分	時間軸 (月)	時間軸 (月)	地域区分	地域区分	/品目分類	548 プリン	【円】
13	1	金額	3	二人以上	0	2E+09	#####	0	全国		132
14	1	金額	3	二人以上	0	2E+09	#####	1003	0110	札幌市	107
15	1	金額	3	二人以上	0	2E+09	#####	2003	0220	青森市	146
16	1	金額	3	二人以上	0	2E+09	#####	3003	0320	盛岡市	109
17	1	金額	3	二人以上	0	2E+09	#####	4003	0410	仙台市	183
18	1	金額	3	二人以上	0	2E+09	#####	5003	0520	秋田市	126
19	1	金額	3	二人以上	0	2E+09	#####	6003	0620	山形市	122
20	1	金額	3	二人以上	0	2E+09	#####	7003	0720	福島市	217
21	1	金額	3	二人以上	0	2E+09	#####	8003	0820	水戸市	147
22	1	金額	3	二人以上	0	2E+09	#####	9003	0920	宇都宮市	182
23	1	金額	3	二人以上	0	2E+09	#####	10003	1020	前橋市	154
24	1	金額	3	二人以上	0	2E+09	#####	11003	1110	さいたま市	182

オープンデータの社会的効果



1. 新産業の創出

- オープンデータを活用した新たなサービスやビジネス、製品の開発
- イノベーションの促進
- 雇用創出



2. コストの節約

- 政府や企業の業務効率化、重複投資の削減
- 公共サービスの最適化
- データ収集コストの低減

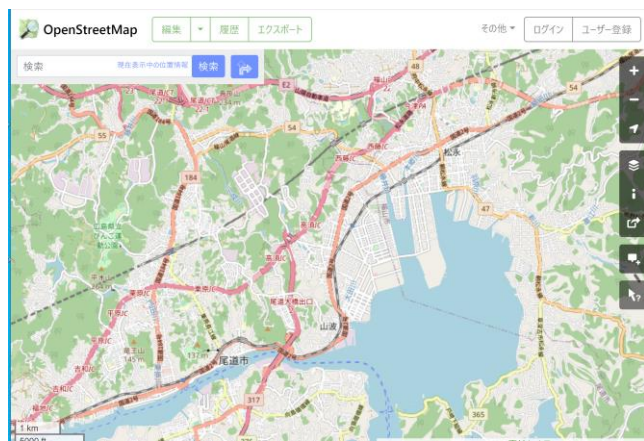


3. 活動の透明性向上

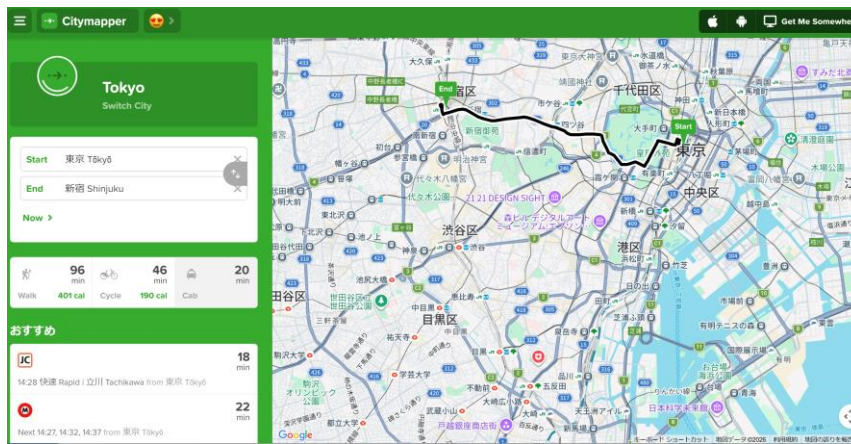
- 行政などの活動が公になり、信頼性や責任感が向上
- 市民の参加促進
- 不正の抑制



オープンデータの例
OpenStreetMap：地図



オープンデータを活用するサービス
Citymapper.com：世界中の大都市の乗り換え案内サービス



オープンデータの活用例：気象庁の気象データ



- 気象庁の気象データは、天気アプリや警報通知として利用されている。
- 利用者は外出や通学の判断をしやすくなり、警報時の早めの安全行動や企業の需要予測にも役立つ。

国土交通省 気象庁 Japan Meteorological Agency

ENGLISH Other Languages 文字サイズ変更 標準 大

気象庁 防災情報 各種データ・資料 地域の情報 知識・解説 各種申請・ご案内

ホーム > 知識・解説 > 急な大雨や雷・竜巻から身を守るために > 気象情報入手するには > 民間気象事業者

民間気象事業者

予報業務許可事業者の携帯端末（携帯電話、スマートフォン等）向け防災情報提供サービスについて

気象庁では、予報業務許可事業者を対象として、携帯端末向け防災情報提供サービスに関するアンケートを行っています。

なお、調査対象のサービスをはじめとした各事業者のサービス内容及び料金等につきましては、それぞれの事業者にお問い合わせ願います。

防災情報提供サービス 実施事業者一覧

五十音順のリストです。事業者名をクリックすると当該事業者の調査結果へ移動します。

あ行				
RSK山陽放送 (株)	いであ(株)	(株)ウェザーマップ	(株)エムティーアイ	(株)応用気象エンジニアリング

現況天気図

天気図 1日(金)午後 6時00分

	2土	3日	4月	5火	6水	7木	8金							
岡山北	晴	晴	晴	晴	晴	晴	晴							
降水	0	0	80	50	20	20	20	30						
岡山南	晴	晴	晴	晴	晴	晴	晴							
降水	0	0	80	50	20	20	20	30						
岡山	晴	晴	晴	晴	晴	晴	晴							
降水	11	25	14	25	13	22	11	23	10	25	12	25	13	25
香川	晴	晴	晴	晴	晴	晴	晴							
降水	0	0	70	50	20	30	20	30						
高松	晴	晴	晴	晴	晴	晴	晴							
降水	16	25	16	25	15	24	13	23	11	24	13	24	14	25

40歳以上なら、このゲームはあなたのためです！
Game of Thrones

台風情報

注意報警報

岡山 香川

天気予報

岡山・香川

今日 明日 週間天気

天気予報

全国

今日 明日 週間天気

気象庁・防災情報提供一覧サイト
<https://www.jma.go.jp/jma/kishou/info/keitai.html>

RKU山陽放送のサイト
<https://www.data.jma.go.jp/developer/weatherdataguide/howto.html>

オープンデータの活用例：GTFS形式の公共交通データ



- GTFS形式の公共交通データにより，時刻表や停留所情報は標準化。乗換案内サービスは使いやすくなる。
- 住民や観光客は，知らない地域でも移動経路や時刻を調べやすくなる。

The screenshot shows the official website of the Ministry of Land, Infrastructure, and Transport (MLIT) of Japan. The page is titled "公共交通政策" (Public Transport Policy) and features a search bar, navigation menu, and a main content area with a cat mascot. The main content area includes a link to "公共交通運行情報標準データ (GTFS-JP) に関する資料・検討会" (Resources and Study Group for Public Transport Operation Information Standard Data (GTFS-JP)).

```
stop_id,stop_code,stop_name,stop_lat,stop_lon,zone_id,location_type
000_2,0,那覇空港国際線,26.207671,127.651151,000_2,0
001_1,1,那覇空港,26.206788,127.651162,001_1,0
001_2,1,那覇空港国内線,26.206395,127.651128,001_2,0
002_1,2,県庁北口,26.213583,127.679819,002_1,0
002_2,2,県庁北口,26.213218,127.679556,002_2,0
003_1,3,泊高橋,26.223624,127.684047,003_1,0
003_2,3,泊高橋,26.224437,127.684967,003_2,0
004_1,4,合同庁舎前,26.227782,127.691933,004_1,0
004_2,4,合同庁舎前,26.227777,127.691584,004_2,0
005_1,5,おもろまち一丁目,26.224896,127.694611,005_1,0
005_2,5,おもろまち一丁目,26.225135,127.694064,005_2,0
006_1,6,おもろまち駅前,26.223456,127.698325,006_1,0
007_1,7,古島駅前,26.232199,127.703992,007_1,0
007_2,7,古島駅前,26.231886,127.704131,007_2,0
008_1,8,大平,26.249437,127.716904,008_1,0
008_2,8,大平,26.248799,127.716555,008_2,0
009_1,9,嘉数,26.255116,127.733333,009_1,0
009_2,9,嘉数,26.254986,127.733072,009_2,0
010_1,10,琉大入口,26.255718,127.763556,010_1,0
010_2,10,琉大入口,26.25547,127.763757,010_2,0
011_1,11,中城,26.27615,127.780749,011_1,0
011_2,11,中城,26.27533,127.780682,011_2,0
012_1,12,喜舎場,26.302308,127.791616,012_1,0
012_2,12,喜舎場,26.302225,127.79203,012_2,0
```

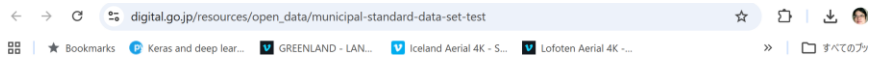
[国土交通省・GTFS 説明サイト](https://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_tk_000067.html)
https://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_tk_000067.html

<https://gtfs-data.jp/search>
で公開されている
同号会社やんばる急行バスの
バス停データ（一部）

オープンデータの活用例：自治体のAED設置場所データ



- 自治体がAED設置場所を標準形式で公開すると、地図やアプリで近くのAEDを探しやすくなる。
- 緊急時の初動を支える救命基盤として機能する。



現在位置： [ホーム](#) > [資料](#) > [オープンデータ](#) > 自治体標準オープンデータセット（正式版）

自治体標準オープンデータセット（正式版）

最終更新日:2025年6月18日

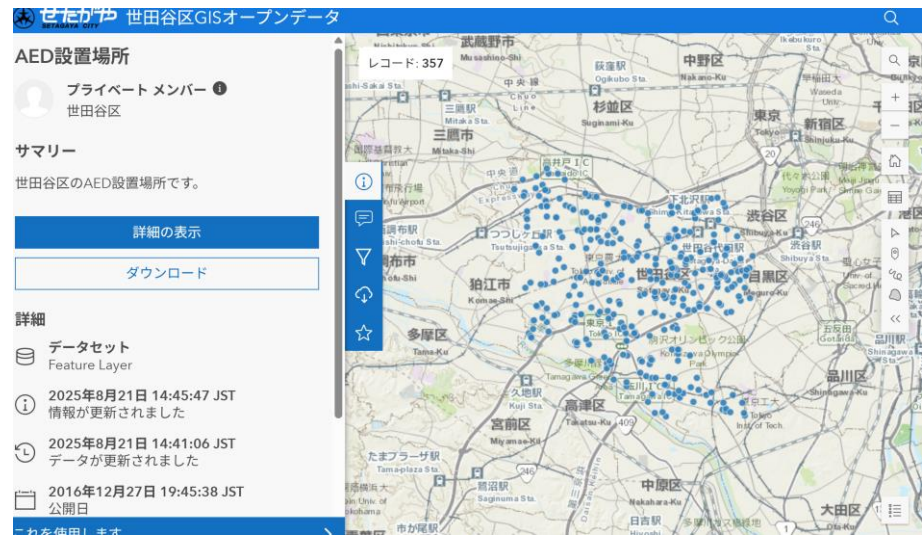
自治体標準オープンデータセット（略称：自治体標準ODS）は、オープンデータの公開とその利活用を促進することを目的としています。公開ニーズの高いデータについて、データ作成時に準拠すべきルールやフォーマット等を取りまとめたものです。

地方公共団体の皆様には、今後オープンデータを更新する際、「自治体標準オープンデータセット」の適用を推奨しています。

2023年3月31日付けで「推奨データセット」から「自治体標準オープンデータセット」へ名称を変更しました。詳細は[解説記事（推奨データセットから自治体標準オープンデータセットへ）](#)（PDF/3,137）をご確認ください。

[デジタル庁の自治体オープンデータセット説明サイト](#)

https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test



世田谷区AEDデータ公開ページ

<https://data-setagaya.opendata.arcgis.com/maps/dc02fd04286954fd9857df355948d1323>

オープンデータの活用例：厚生労働省の感染症動向データ



- 感染症動向データにより，流行状況が把握でき，住民や医療機関の判断材料になる。
- 受診や感染対策，事前の備えを考える際に役立つ。



感染症発生動向調査

感染症発生動向調査とは

感染症発生動向調査は、昭和56年から開始され、平成11年4月に「感染症の予防及び感染症の患者に対する医療に関する法律」（平成10年法律第114号。以下「感染症法」という。）が施行されたことに伴い、感染症法に基づく施策として位置づけられた調査です。

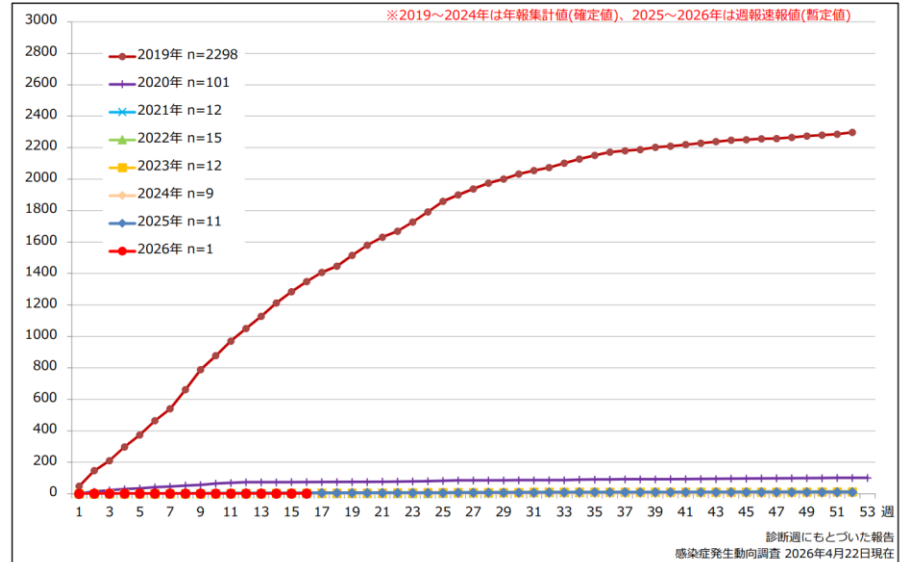
感染症の発生情報の正確な把握と分析、その結果の国民や医療機関への迅速な提供・公開により、感染症に対する有効かつ的確な予防・診断・治療に係る対策を回り、多様な感染症の発生及びまん延を防止することを目的としています。

対象感染症

- 政策について
- 分野別の政策一覧
- 健康・医療
 - 健康
 - 食品
 - 医療
 - 医療保険
 - 医薬品・医療機器

1. 風しん累積報告数の推移 2019～2026年（第1～16週）

Cumulative rubella cases by week, 2019-2026 (week 1-16) (based on diagnosed week as of April 22, 2026)



風しん発生状況（一部）

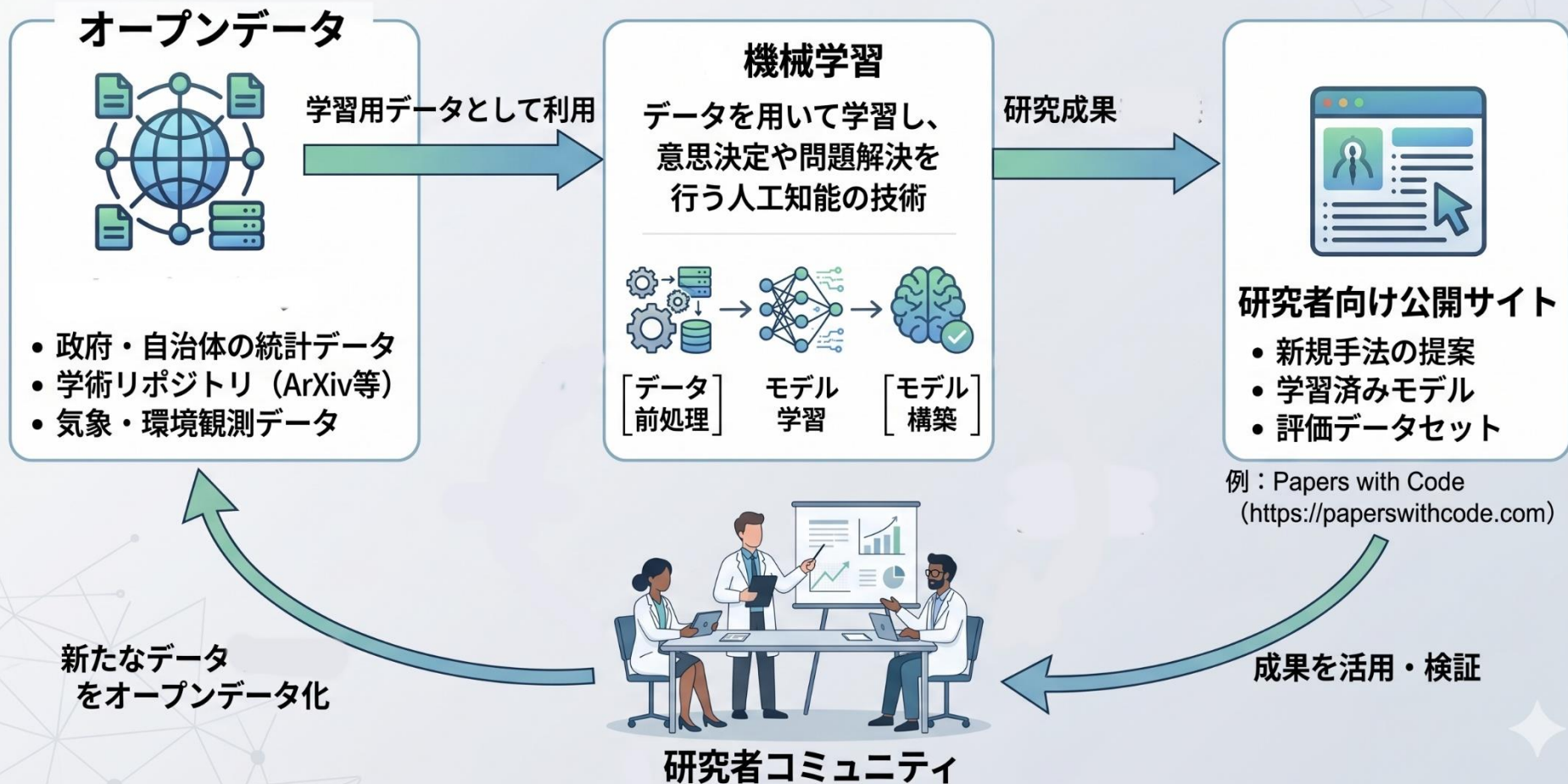
<https://id->

[info.jihs.go.jp/surveillance/idwr/diseases/rubella/graph/2026/rube26-16.pdf](https://id-info.jihs.go.jp/surveillance/idwr/diseases/rubella/graph/2026/rube26-16.pdf)

[厚生労働省・感染症発生動向調査の説明ページ](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000115283.html)

<https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000115283.html>

オープンデータと機械学習との関係



オープンデータ利用上の注意点



利用者側が注意深く利用する必要がある

確認事項

- データの品質や信頼性
- プライバシーの侵害をしていないこと

遵守事項

- 著作権の尊重
- 出典の明示
- 公開者が定める利用条件の遵守

「公開されているデータだから100%正確、何をしてもよい」というわけではない

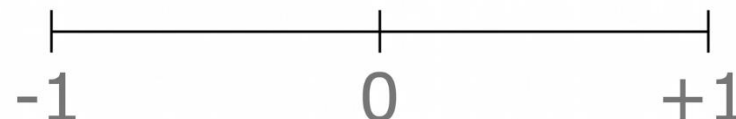
3-2 クロス集計表と相関係数

クロス集計表

カテゴリデータ（性別，居住地など）の関係分析に用いる

相関係数

数値データ（金額，個数など）の関係分析に用いる



共通点

ともに『2つの変数の関係进行分析する手法』であるが，対象データの種類が異なる

クロス集計表

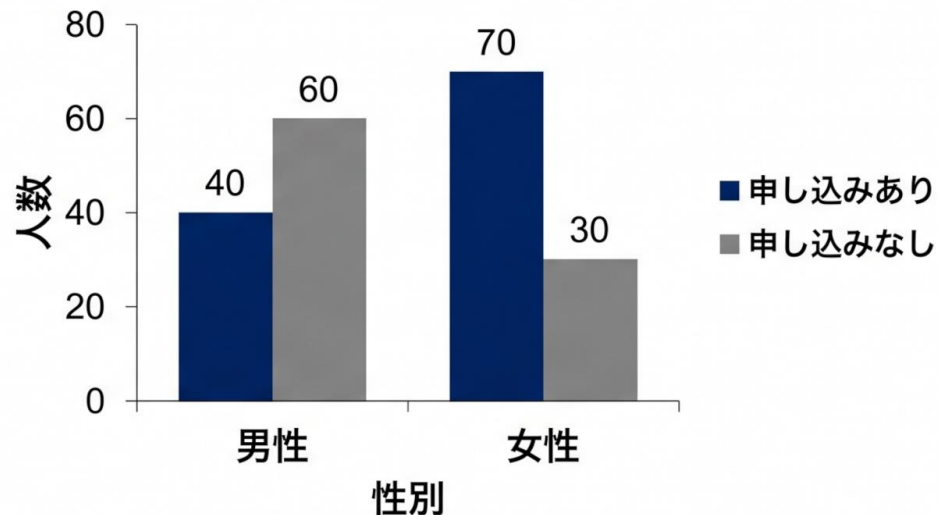


- 2つの変数の関係を分析
- カテゴリデータ用（年代、種類、等級、成否など）

クロス集計表の例：性別と申し込み状況

	申し込みあり	申し込みなし
男性	40	60
女性	70	30

棒グラフによる可視化（同じデータを使用）



分析の目的

性別と申し込み状況に関連があるかを分析する

クロス集計表と元データ



元データ（回答者200名）

No.	性別	申し込み状況
1	男性	申し込みあり
2	女性	申し込みなし
3	男性	申し込みなし
4	女性	申し込みあり
5	男性	申し込みあり
...
200	女性	申し込みあり

合計 200名

集計



200名の回答を
性別×申し込み
状況で集計

クロス集計表：性別と申し込み状況

	申し込みあり	申し込みなし
男性	40	60
女性	70	30

合計 $40 + 60 + 70 + 30 = 200$ 名

- クロス集計表は、2つの変数（例：性別×申し込み状況）の組み合わせごとに人数等を集計した表である。
- 元データを2変数で分類することで、変数間の関連（例：女性の申し込み率が高い傾向）を把握できる。

クロス集計表を用いた分析例

最終卒業学校卒業のあと、引っ越しをしていないのは全体の何パーセントか？

最終卒業学校卒業時の居住地（男女別）

	まだ卒業していない	現在と同じ居住地	その他
男性	19.2	17.8	63.0
女性	17.3	10.6	72.1

パーセント

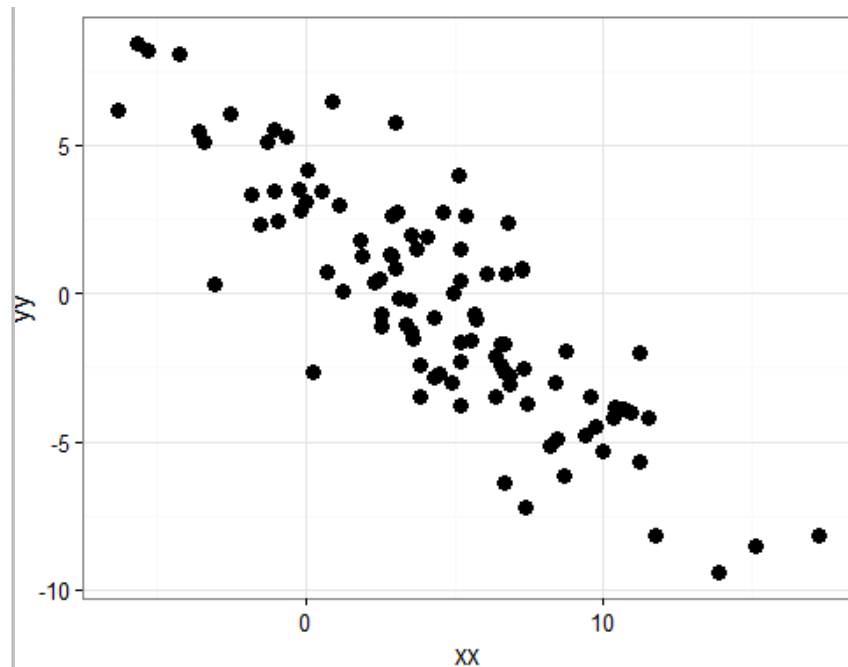
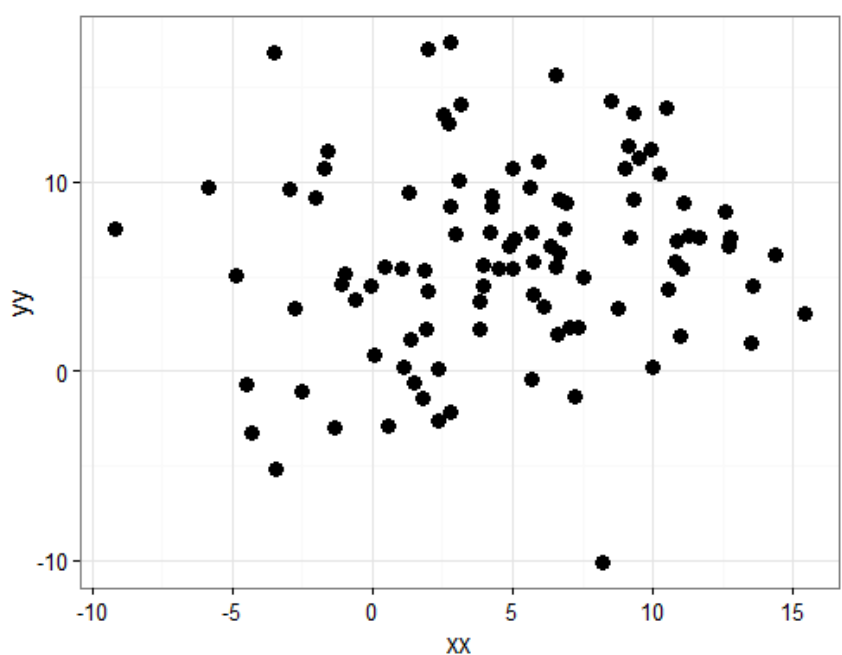
ここから導かれる結論は？
性別と居住地についての関連

その他：
居住地が違う，不詳

相関係数を求める前に散布図で確認



- **散布図**：2つの変数の関係を**点の集まり**で視覚化する
- **構成**：**横軸に一方の変数**，**縦軸にもう一方の変数**を取り，データを点としてプロット
- **推奨**：深い分析の前に，**まず散布図などでデータの分布や関係性の傾向を視覚的に確認**することが望ましい



相関の定義と種類 (正・負・なし)

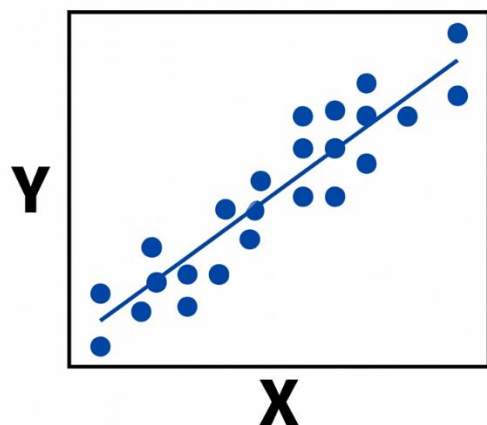


2つの変数の関係进行分析 (数値データ用)

相関がある場合，一方が変化するともう一方も変化する傾向にある

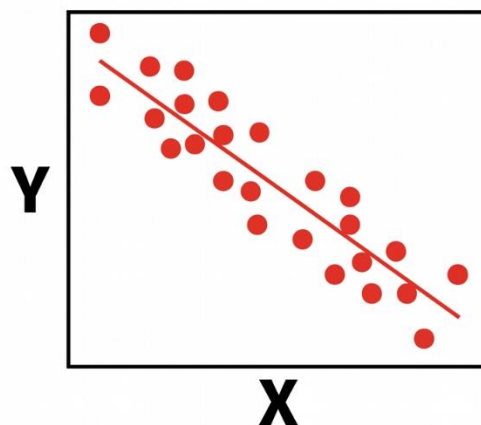
種類

正の相関



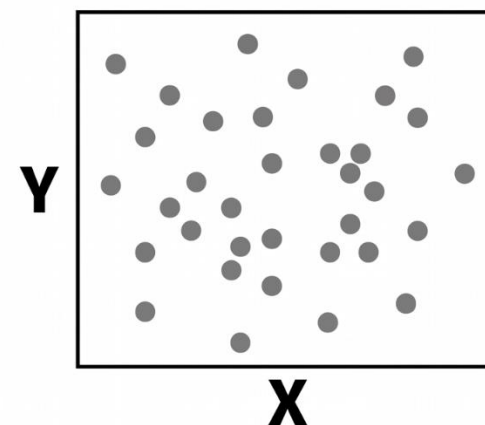
Xが増えるとYが増える傾向
例：勉強時間が増えると
得点上がる

負の相関



Xが増えるとYが減る傾向
例：ガソリン代が上がると
車の利用が減る

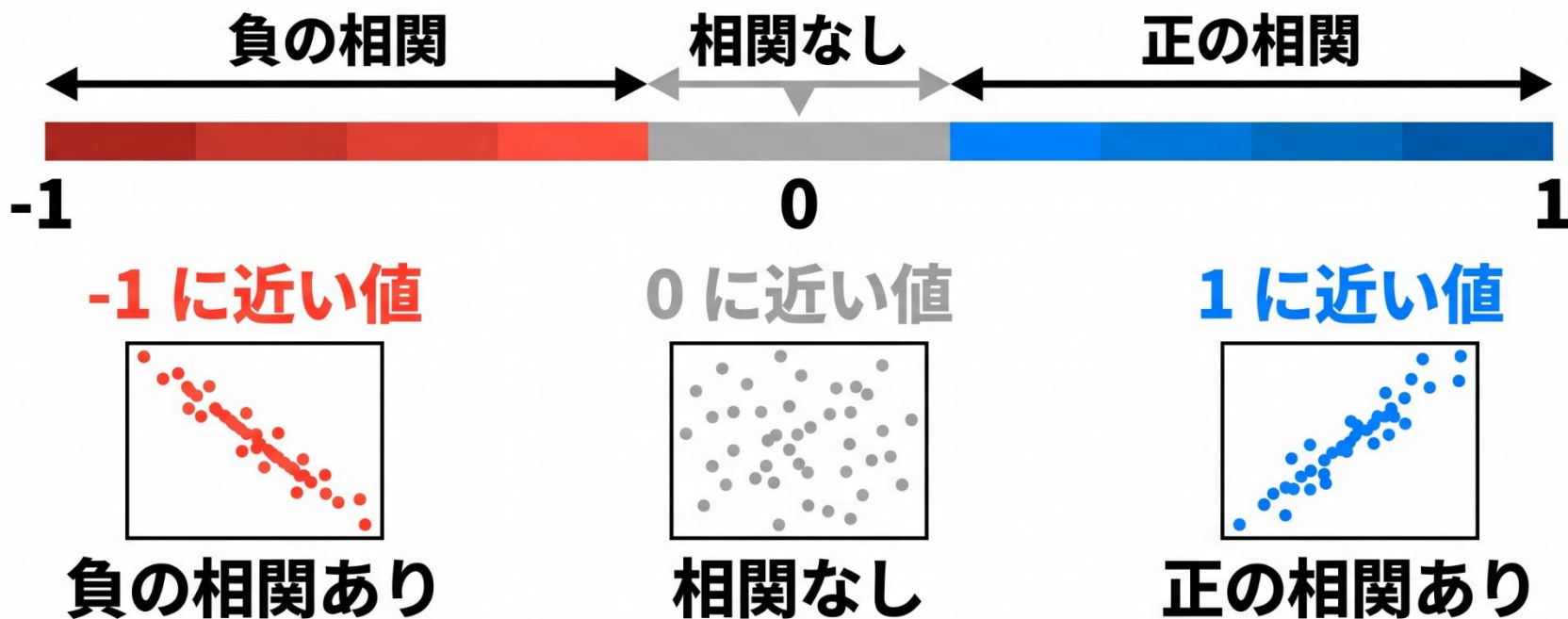
相関なし



XとYに関係がない
例：足のサイズと勉強時
間に関係がない

相関係数：相関を算出した数値

範囲：-1 から 1 まで

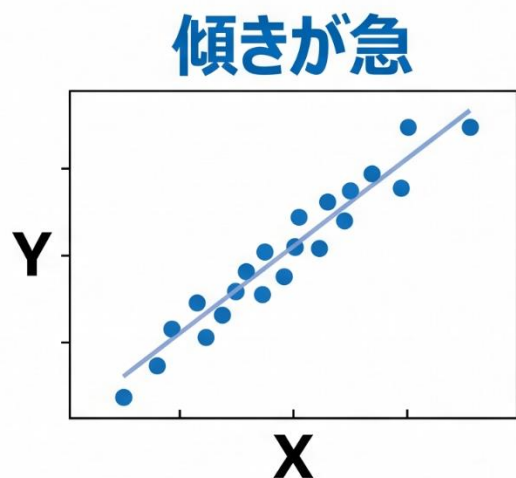


相関係数を算出することで、変数間の関係の分析ができる

相関係数の性質

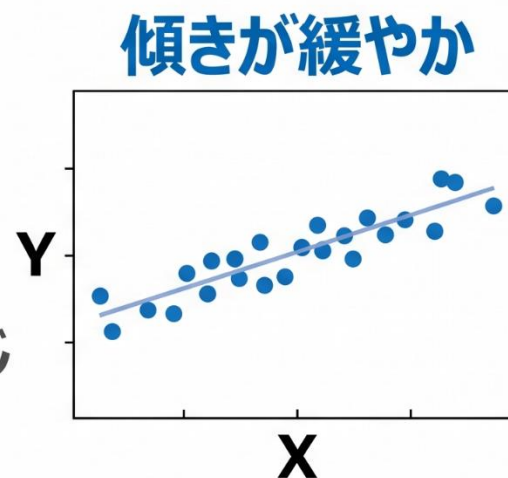


相関係数は『相関の強弱』の尺度であり、『傾き』ではない
データの散らばり具合がそろっていれば、傾きが緩やかであっても1に近い値を示す



相関係数 $\doteq 1$
強い正の相関

=
相関の強さは同じ



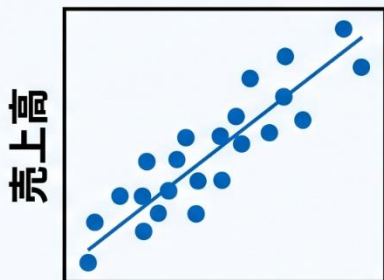
相関係数 $\doteq 1$
こちらも強い正の相関

相関係数が示すのは『点のばらつき小ささ』であり、直線の傾きの大きさではない



2つの量の間の関係性の分析に活用される

① マーケティング

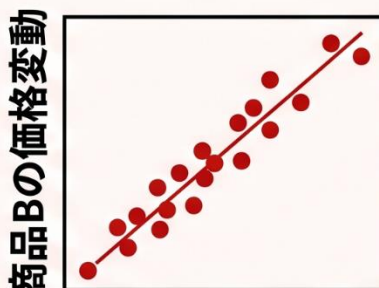


広告費

相関係数 $\doteq 0.8$

広告を増やすと売上高
が増えそうか

② 金融



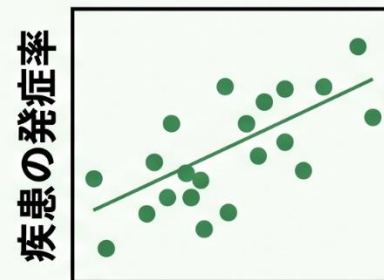
商品Bの価格変動

商品Aの価格変動

相関係数 $\doteq 0.9$

相関が高い複数の金融商
品を扱うとリスクが高いか

③ 医療・生命科学



疾患の発症率

遺伝子の発現量

相関係数 $\doteq 0.6$

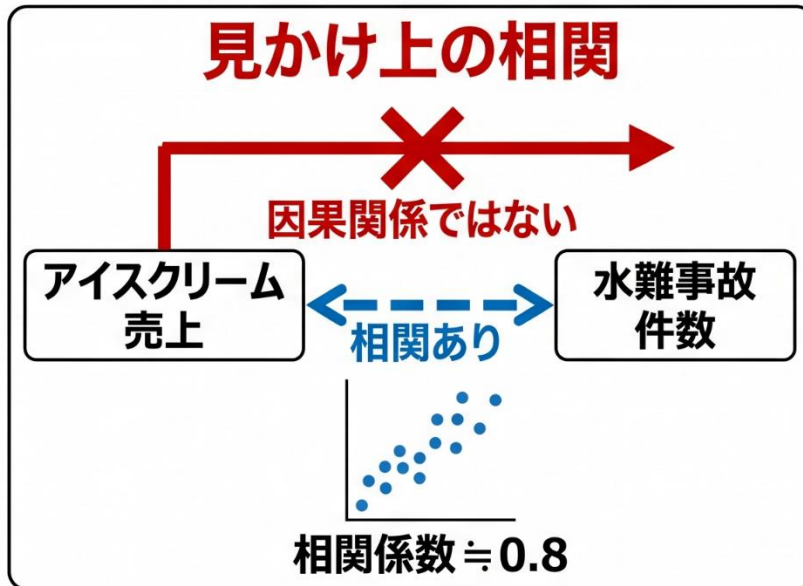
遺伝子と疾患に
関係がありそうか

いずれも2つの量の間の関係性の分析に活用される

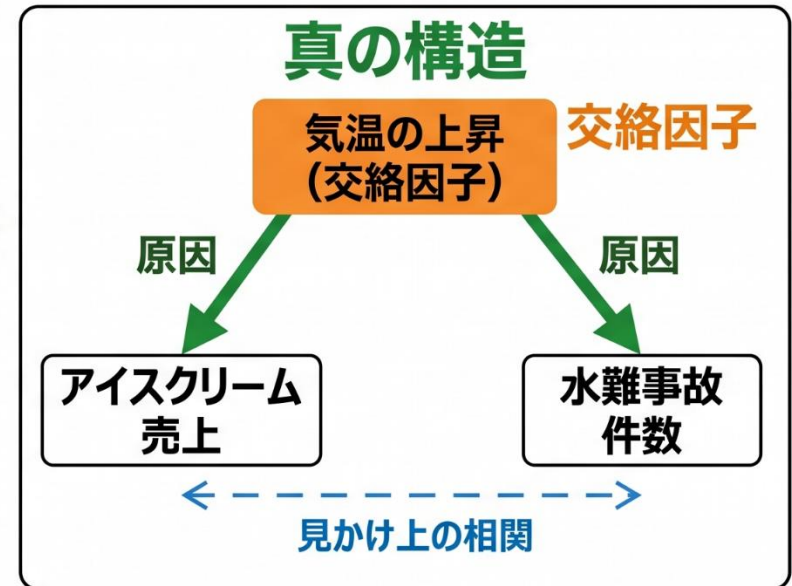
相関と因果関係の違い



相関があっても因果関係があるとは限らない
両変数に共通して影響を与える第三の変数（**交絡因子**）が存在する場合、
見かけ上の相関が生じる



正しく
分析すると



データ分析では『相関=因果』と短絡しないことが重要である

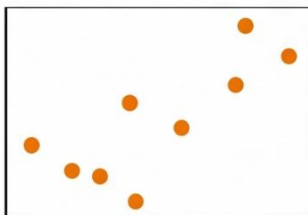
正確な判断には、サンプル数やデータの偏りの確認が必要 (数字だけ見ての判断は危険)



実データには測定誤差やばらつきが含まれる

サンプル数が少ない場合や偶然による偏りがある場合、相関係数の値だけで結論を下すことは危険である

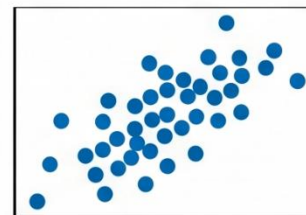
サンプル数が少ない



相関係数 ≈ 0.8 (偶然の可能性)

少数データでは偶然の偏りが相関係数を大きくすることがある

サンプル数が十分



相関係数 ≈ 0.5 (信頼できる推定)

十分なデータ数で相関の強さを正しく評価

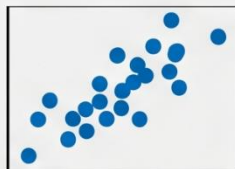


サンプル数で信頼性が変わる

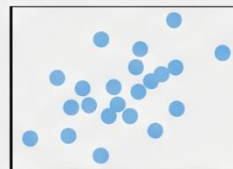
相関係数の目安 (あくまで目安)



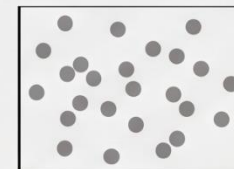
0.7 以上
強い相関



0.4 ~ 0.7
中程度



0.2 ~ 0.4
弱い相関

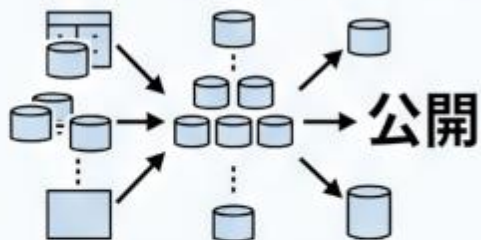


0.2 未満
相関なし

全体まとめとデータの適切な取り扱い



① オープンデータ



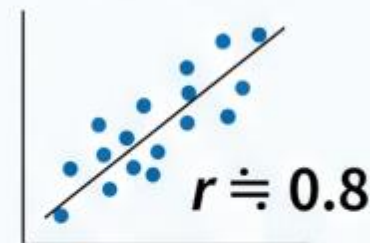
様々な分野のデータを網羅的に公開
社会課題解決・効率化・透明性向上

② クロス集計表

	12	8
	5	15

質的変数（カテゴリ）間の関係を表形式で整理

③ 相関係数



量的変数（数値）間の関係の強さを数値化

① プライバシー保護

個人を特定できる情報の適切な管理

② 個人情報の取扱い

本人の同意なき目的外利用の禁止

③ AI倫理

偏りや誤りを含むデータはモデルの出力に影響する

④ 情報漏えいの防止

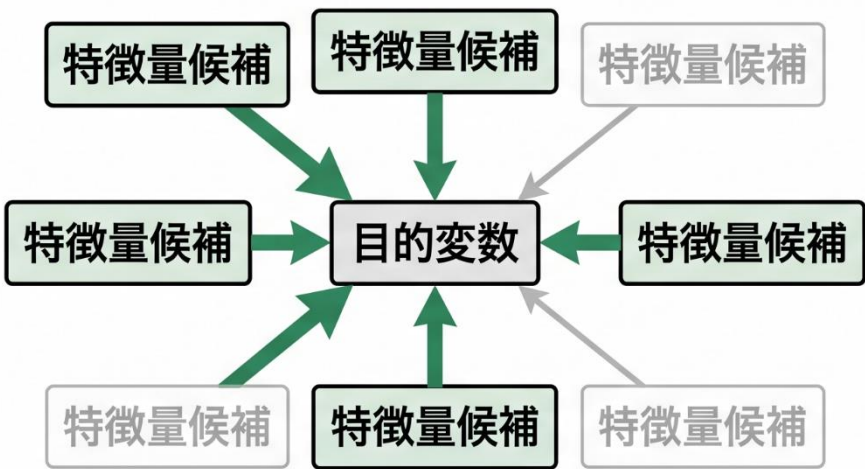
アクセス管理と適切な保管・通信

「今回の学び」のAIへの接続



機械学習における特徴量の把握

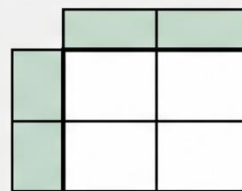
大量のデータの中から、予測に役立つ変数（特徴量）を把握することが重要



関連性の強い特徴量を選ぶと予測精度が高まる

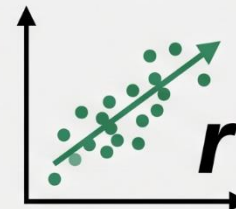
関連性の把握に活用できる手法

クロス集計表



変数間の関連性を把握

相関係数



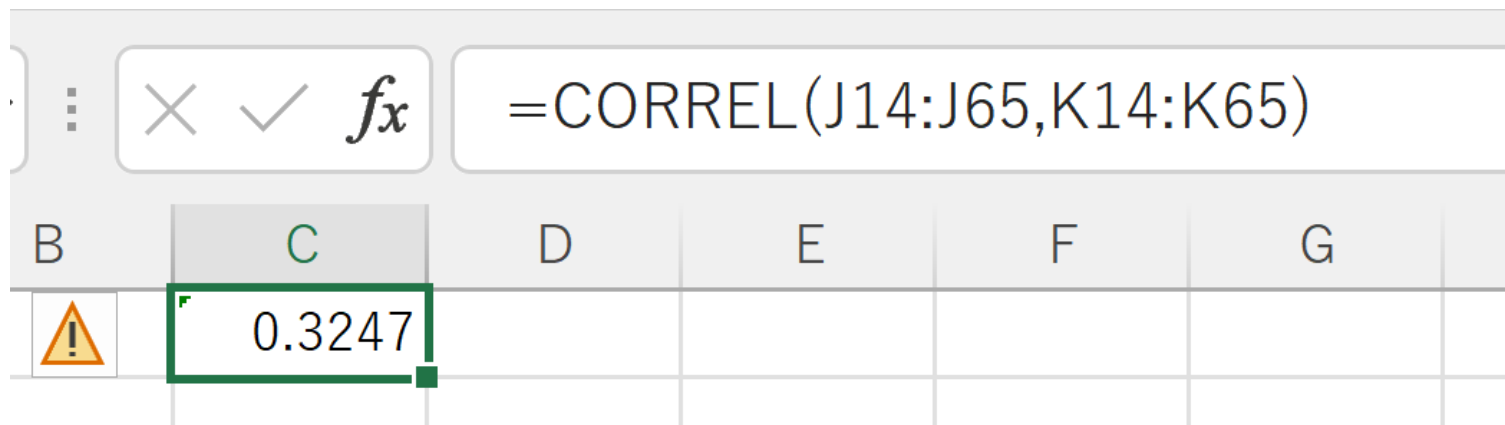
変数間の関連性の強さを数値で把握

特徴量の選択に活用

データを正しく理解し適切に扱うことが、分析とAI活用の基盤となる

Excel で相関係数を求める

- ① 相関係数を求めたいセルを選んでおく
- ② メニューの「**数式**」
- ③ 「**その他の関数**」, 「**統計**」, 「**CORREL**」
- ④ **配列 1 と配列 2 の範囲**を指定



相関係数として 0.3247 が求まった

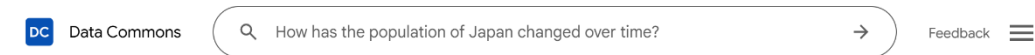
演習



演習① 質問を理解するAI 「Data Commons」



- **オープンデータ**を、AIの力を借りて体験
- Data Commonsは、**Googleが世界中の政府統計を1か所に集めたサービス**
- **検索欄には英語で質問**を入力する。英語が苦手な場合は、ホームページ資料の質問例をそのままコピーして貼り付ければよい。



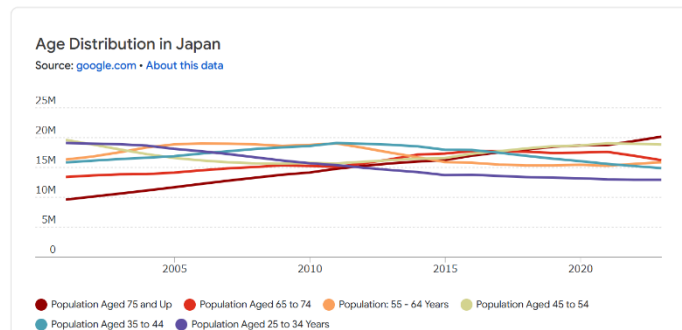
About [Japan](#), country in [Asia](#) • Knowledge Graph [country/JPN](#)

Data related to your research question

How has the population of Japan changed over time?

Age Distribution

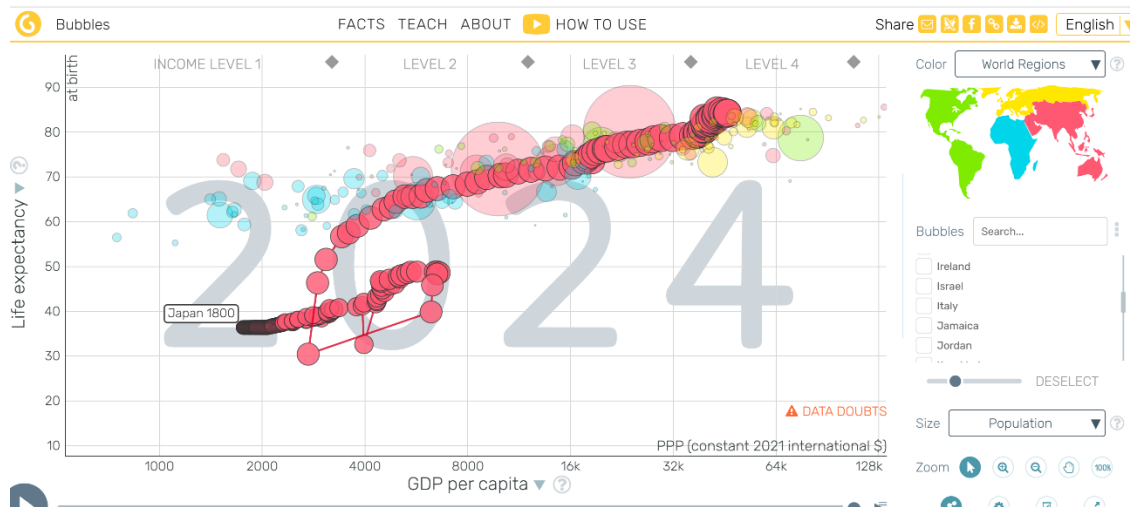
One facet available for this chart See per capita



演習② データの隙間を埋めるAI 「Gapminder」



- オープンデータを体験
- Gapminderは、**200年分の世界の統計を動くバブルで見せる**サイト
- スウェーデンの医師ハンス・ロスリングらが、**世界の貧困や健康の変化を一目で分かるように**するために作った。
- バブルの**大きさ**は**各国の人口**を、バブルの**色**は**地域**（アジア・ヨーロッパ等）を表す。日本のバブルがどのように動くかを、自分の目で追ってほしい。

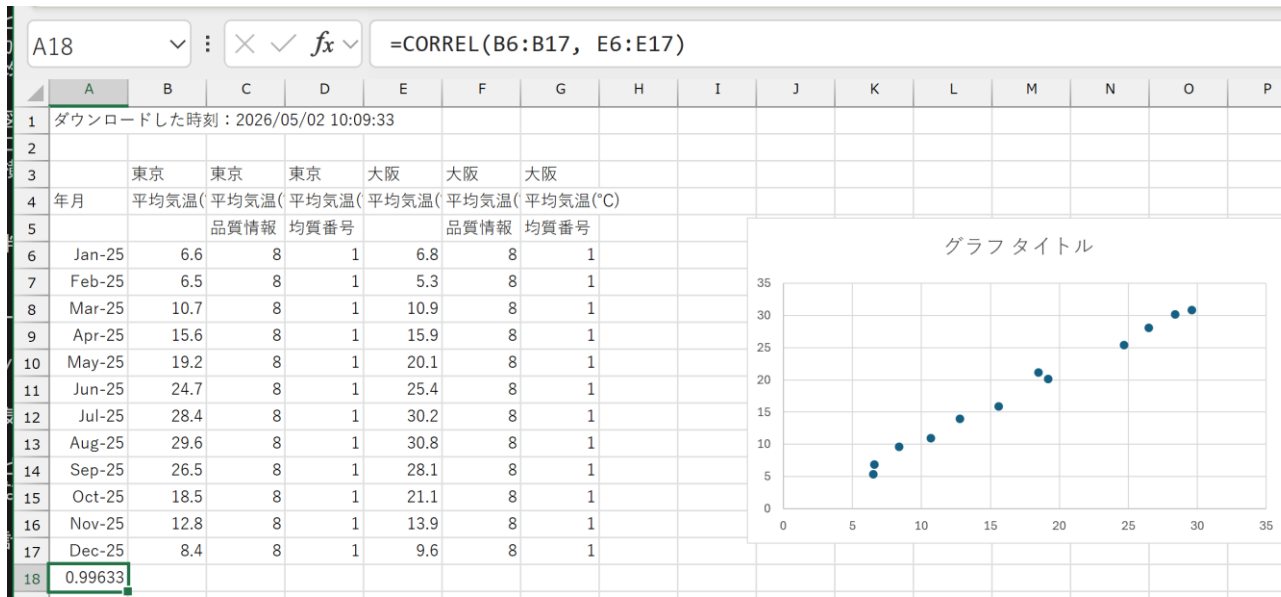


演習③ 過去の気象データ・ダウンロードと相関の分析



- オープンデータから自分で相関係数を求め、数値をもとに判断を下す体験
- 2つの観測地点を選んで気温や降水量をCSVで取り出し、Excelで相関係数を計算する。

(例) 「東京と大阪の月平均気温に相関はあるか」を調べる。
地理的に離れていても日本国内の2地点なら気温は連動するはずである。自分の目で数値的に確かめてほしい。



演習④ 総務省統計局「統計でみる都道府県のすがた」から相関の分析



- オープンデータから自分で相関係数を求め、数値をもとに判断を下す体験
 - 総務省統計局の「統計でみる都道府県のすがた」
 - 47都道府県ごとに、人口・産業・教育・医療・住宅など多数の指標を1つのExcelファイル（または分野ごとのファイル）にまとめて公開している。
 - ファイルを開けばすぐに相関係数の計算に使える。
- (例) 「人口と医師数に相関はあるか」を調べる。

	G	H	I	J	L	M	N	O
3								
4					A. 人口・世帯		A Population and Household	
6					No. 1		No. 2	No
7					総人口		総人口(男)	
8					Total population		Total population (Male)	
9					#A011000		#A0110001	#A
10		単位	Unit		万人：10 thousand persons		万人：10 thousand persons	
11		年度	Fiscal year		2024		2024	
12		都道府県	Prefecture		指標値	順位	指標値	順位
					Indicator	Rank	Indicator	Rank
66								

演習⑤ e-Stat「統計ダッシュボード」から相関の分析



- オープンデータから**自分で相関係数を求め**、数値をもとに**判断を下す**体験
- e-Statの「**統計ダッシュボード**」
- 指標ごとに都道府県のランキングが画面上で確認できる。
- 演習④とは別の指標を選び、画面に表示された値をExcelに取り込んで、グラフを作成したり、相関係数を求めたりしてみる

The screenshot shows the 'Stat Dashboard' interface for population data. The main content is a table with columns for '地域別' (Regional), '総人口 (総数) 【人】' (Total Population), and '日本人人口 (総数) 【人】' (Japanese Population). The table lists data for various regions, with Aomori Prefecture highlighted in blue. The interface includes navigation menus for 'データを見る' (View Data), '時系列表' (Time Series Table), and 'ヘルプ' (Help). A sidebar on the left allows for filtering by '地域階級' (Regional Hierarchy) and '地域別' (Regional).

地域別	総人口 (総数) 【人】	日本人人口 (総数) 【人】
北海道	5,043,000	4,980,000
青森県	1,165,000	1,157,000
岩手県	1,145,000	1,134,000
宮城県	2,248,000	2,219,000
秋田県	897,000	891,000
山形県	1,011,000	1,001,000
福島県	1,743,000	1,724,000
茨城県	2,806,000	2,716,000
栃木県	1,885,000	1,830,000
群馬県	1,890,000	1,811,000
埼玉県	7,332,000	7,083,000
千葉県	6,251,000	6,057,000
東京都	14,178,000	13,463,000
神奈川県	9,225,000	8,941,000