

# ビジネスパーソン向け 統計データ活用セミナー

跡見学園女子大学  
山澤成康

# プログラム概要

- 1. なぜ今「データ分析×政府統計」か？
- 2. データ分析は何から始めればいいのか？
- 3. Excelを使った政府統計データの分析（演習）

# 1. なぜ今「データ分析×政府統計」か？

- ① データ分析に政府統計を使うメリット
  - 企業内データだけでは見えない社会的傾向を補完
- ② ビジネスシーンでの活用可能性
  - 市場分析・需要予測・地域比較などへの応用
- ③ 政府統計データ活用の先行事例紹介
  - 民間企業での具体的導入例を紹介

- ①データ分析に政府統計を使うメリット  
ー 企業内データだけでは見えない社会的傾向を補完

# 政府統計のメリット

## 信頼性

公的に整備された客観データ  
経営計画・報告書

## 多様なデータ

市場・需要の把握  
新商品開発・販路拡大

## マクロ経済の把握

景気・産業の動向分析  
リスクマネジメント

## 詳細さ

立地・商圈分析  
店舗展開計画

## コスト削減

無料の高品質データ  
分析の基本

## 説得力

公的根拠で説得力UP  
提案書・IR資料

- ①データ分析に政府統計を使うメリット  
 - 企業内データだけでは見えない社会的傾向を補完

1. なぜ今「データ分析 × 政府統計」か？

2. データ分析は何から始めればいいのか？

3. Excelを使った政府統計データの分析(演習)

# 主な統計の紹介

	統計名称	調査項目
人口統計	国勢調査	年齢、性別、世帯構成、職業など
		世帯数、家族構成、住宅情報
	人口推計	年齢別人口、男女別人口、地域別人口
生活関連統計	家計調査	消費支出、収入、貯蓄・負債
	労働力調査	就業者数、就業形態、失業率
産業・企業統計	経済センサス	企業数、業種別売上高、従業員数
	鉱工業指数	製造業や鉱業の生産数量、出荷額、在庫水準
	商業動態統計調査	小売業、卸売業の売上高、在庫、販売数量

- ①データ分析に政府統計を使うメリット  
ー 企業内データだけでは見えない社会的傾向を補完

## 統計の紹介

### 家計調査

- 毎月何にどの程度支出したか
- 収入階級別、年齢別

### 国勢調査

- 地域ごとの人口、世帯

### 経済センサス 活動調査

- 地域ごとの事業所数、従業員数、売上高

- ①データ分析に政府統計を使うメリット  
 - 企業内データだけでは見えない社会的傾向を補完

1. なぜ今「データ分析 × 政府統計」か？

2. データ分析は何から始めればいいのか？

3. Excelを使った政府統計データの分析(演習)

# ビジネスシーンでの活用例 (1)

分野	活用内容	主な利用統計例
① 市場分析	商品の特性を分析する	家計調査
② 地域分析・立地戦略	地域ごとの人口構成、所得水準、産業構造を分析して出店・投資先を決定	国勢調査、経済センサス活動調査
③ 顧客セグメンテーション	年齢・性別・職業・家族構成などの人口統計からターゲット顧客層を設定	国勢調査、就業構造基本調査
④ 景気・業界動向分析	マクロ経済・産業別動向をモニタリングし、経営判断や在庫調整に活用	景気動向指数、鉱工業生産指数、法人企業統計調査
⑤ 労働市場・人材戦略	労働需給や賃金動向を分析し、採用・給与戦略に反映	労働力調査、賃金構造基本統計調査

- ①データ分析に政府統計を使うメリット  
 - 企業内データだけでは見えない社会的傾向を補完

1. なぜ今「データ分析  
×政府統計」か？

2. データ分析は何から  
始めればいいのか？

3. Excelを使った  
政府統計データの  
分析(演習)

## ビジネスシーンでの活用例 (2)

分野	活用内容	主な利用統計例
⑥サステナビリティ・ESG対応	CO <sub>2</sub> 排出、エネルギー消費、廃棄物量などの統計で環境経営を支援	総合エネルギー統計
⑦観光・インバウンド分析	訪日外国人の動向や旅行支出の分析で観光ビジネスを最適化	インバウンド消費動向調査 (旧 訪日外国人消費動向調査)
⑧不動産・都市開発分析	地価・住宅需要・建築動向を把握し、開発計画や投資判断に活用	住宅・土地統計調査、建築着工統計調査
⑨リスクマネジメント・事業継続計画 (BCP)	災害リスク、人口減少、感染症動向などを定量的に評価	気象統計、人口動態調査
⑩イノベーション・新規事業探索	統計を組み合わせて、未開拓ニーズ・成長分野を特定	科学技術研究調査

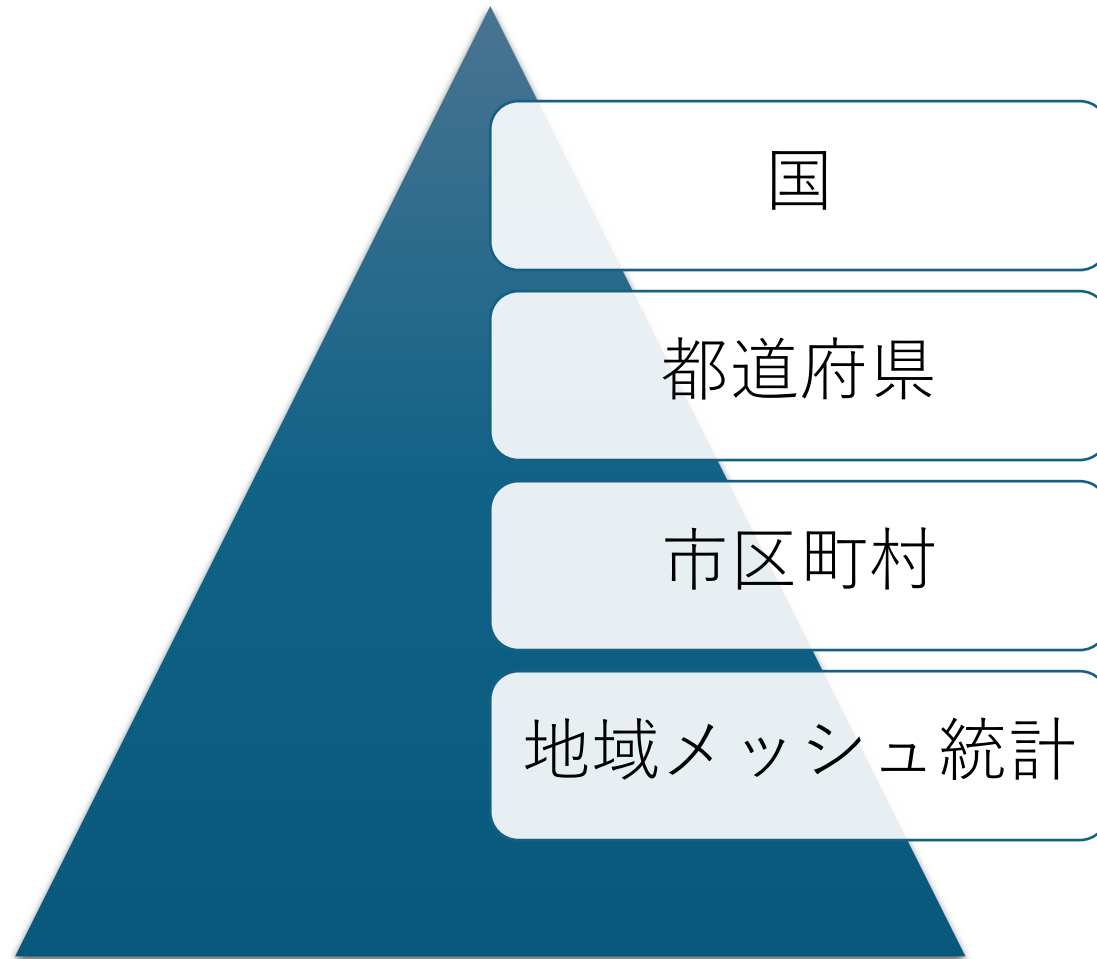
- ①データ分析に政府統計を使うメリット  
ー 企業内データだけでは見えない社会的傾向を補完

1. なぜ今「データ分析  
×政府統計」か？

2. データ分析は何から  
始めればいいのか？

3. Excelを使った  
政府統計データの  
分析(演習)

# 地域別データで地域特性も可視化が可能



GDP統計、鉱工業生産  
指数など

「統計でみる都道府県の  
すがた」参照

「統計でみる市区町村  
のすがた」参照

①データ分析に政府統計を使うメリット  
 - 企業内データだけでは見えない社会的傾向を補完  
 参考：「統計でみる都道府県のすがた」  
 「統計でみる市区町村のすがた」

「統計でみる都道府県のすがた」「市町村のすがた」は、社会・人口統計体系の指標データの中から自治体別の主要なものを抜粋し報告書に取りまとめたもの。数値を見やすくするための目安として、自治体別順位を付している。

総合統計書のご案内

統計でみる都道府県のすがたの概要

- 統計の目的
- 統計の沿革

本書の内容

- 利用上の注意
- 統計表
- バックナンバー
- 正誤情報

公表予定

Q&A

問合せ先

統計でみる都道府県のすがた

統計でみる都道府県のすがた 2025

Statistical Observations of Prefectures

「統計でみる都道府県のすがた」は、社会・人口統計体系の指標データの中から、都道府県別の主要なものを抜粋し取りまとめたもので、408指標を掲載しています。[2025年2月刊行]  
 本書のほか「統計でみる市区町村のすがた」も刊行しています。



更新情報

- 2025年2月21日 統計でみる都道府県のすがた 2025を掲載しました。

データ表示 (都道府県データ) 市区町村データへ切り替え

統計表表示 | グラフ表示

地域選択 | 表示項目選択 | 統計表表示

地域選択 ?

任意の地域 | 類似の地域

地域を選択してください。

1 絞り込み

地域区分  都道府県  地方 類型 | 地方の選び方

2 地域候補

地域候補 48地域

00000\_全国  
 01000\_北海道  
 02000\_青森県  
 03000\_岩手県  
 04000\_宮城県  
 05000\_秋田県  
 06000\_山形県  
 07000\_福島県  
 08000\_茨城県  
 09000\_栃木県  
 10000\_群馬県

地域を選択>  
 全て選択>  
 <地域を削除  
 <全て削除

選択中地域 0地域 (残り 2500 地域選択可能) クリア

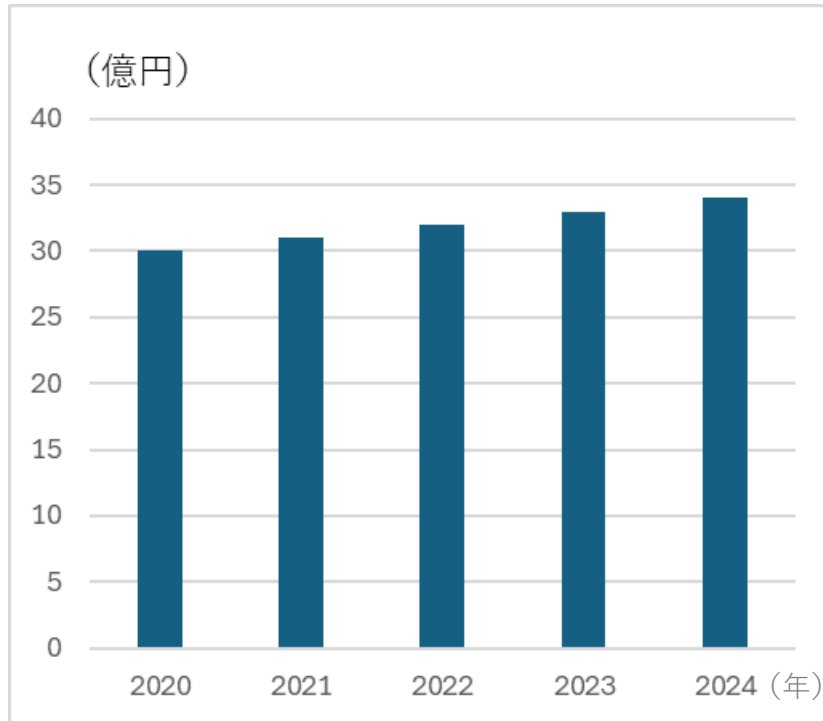
キャンセル 確定

URL : <https://www.stat.go.jp/data/k-sugata/index.html>

- ①データ分析に政府統計を使うメリット  
ー 企業内データだけでは見えない社会的傾向を補完

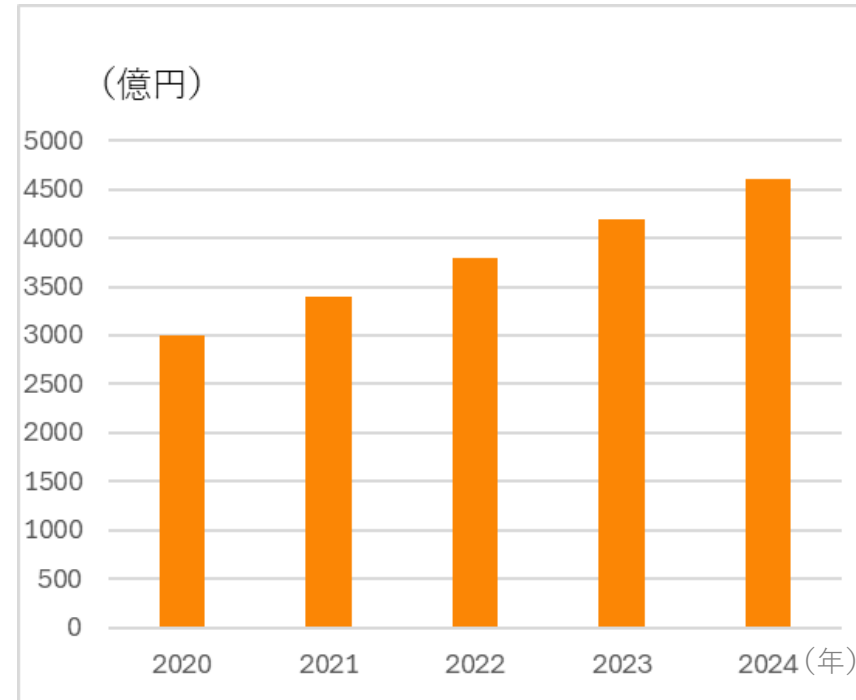
# 企業の状態を社会全体の傾向と比較

自社の製品が伸びていても…



A社の外食部門の売上高

国全体ではもっと伸びている場合も



国全体の外食の消費額 (仮設値)

- ② ビジネスシーンでの活用可能性  
ー 市場分析・需要予測・地域比較などへの応用

## マーケティングツールとしての家計調査

- 「ビッグデータ」
- さまざまな商品の特性を捉えられる
- 日本全体の傾向がわかる
- 総額、収入階級別、年齢階級別で分析
- 金額、数量、価格面からも分析可能（今回は金額中心）
- 高い信頼性

## ② ビジネスシーンでの活用可能性

ー 市場分析・需要予測・地域比較などへの応用

# 家計調査で2010年に比べて伸び率が高い品目

	品目分類 (2020年改定)	2010年	2024年	伸び率(%)
1	766 携帯電話機【円】	2,392	8,037	236.0
2	75B レンタカー・カーシェアリング料金【円】	1,236	4,122	233.5
3	953 寄付金【円】	3,797	10,426	174.6
4	3X5 ウイスキー【円】	1,252	2,887	130.6
5	38A 乳飲料【円】	1,229	2,830	130.3
6	353 チョコレート菓子【円】	1,143	2,568	124.7
7	363 他の主食的調理食品【円】	9,056	20,033	121.2
8	316 キウイフルーツ【円】	1,073	2,352	119.2
9	365 サラダ【円】	2,936	6,351	116.3
10	387 炭酸飲料【円】	3,666	7,722	110.6
11	1.6.2 果物加工品【円】	1,830	3,745	104.6
12	370 冷凍調理食品【円】	5,432	11,032	103.1
13	676 子供用靴・サンダル(2019年までは運動靴、サンダルを含	1,132	2,270	100.5
14	38B ミネラルウォーター【円】	2,362	4,558	93.0
15	229 他の加工肉【円】	1,793	3,452	92.5
16	397 喫茶代【円】	5,143	9,837	91.3
17	711 保健用消耗品【円】	6,727	12,738	89.4
18	906 歯磨き【円】	2,090	3,936	88.3
19	361 調理パン【円】	3,731	6,915	85.3
20	299 他の野菜・海藻加工品のその他【円】	1,361	2,460	80.7

## 家計調査 2人以上の世帯 年次 品目分類 (2020年改定) (総数・金額)

2人以上世帯の約600品目の消費額について、伸び率の高い順に並べ替えたもの。

結果表に記載された2010年～2024年の伸び率を詳しく把握したい場合は、グラフ化して見える化することも可能。

(→次頁以降参照)

(出典) 家計調査

② ビジネスシーンでの活用可能性  
ー 市場分析・需要予測・地域比較などへの応用

# レンタカー・カーシェアリング支出額 (2010～2024年経年)

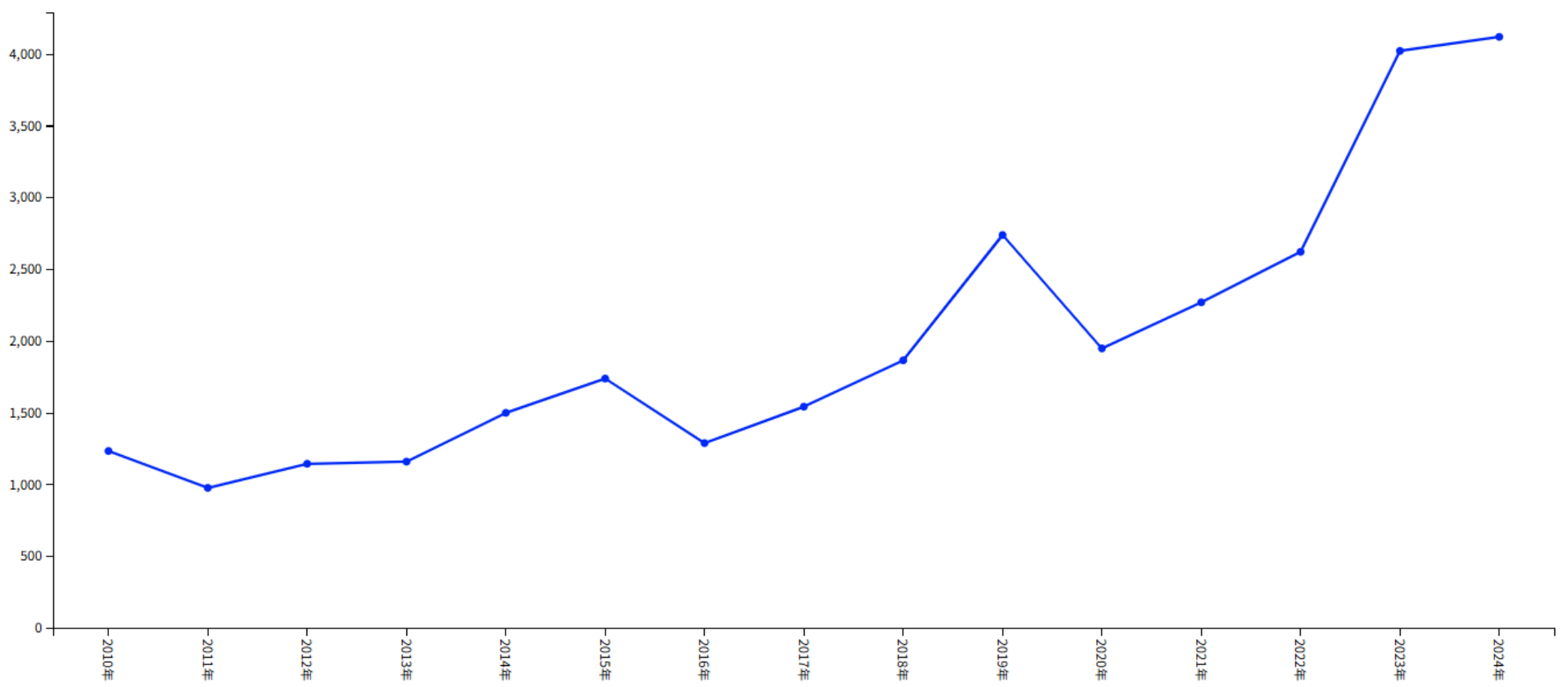
▼ もっと見る

統計表表示

グラフ表示



表章項目  世帯区分 (年次一人以上の世帯)  地域区分



■ 75B レンタカー...

表示項目選択 >

レイアウト設定 >

グラフ表示設定 >

② ビジネスシーンでの活用可能性  
ー 市場分析・需要予測・地域比較などへの応用

# チョコレート菓子支出額（2010～2024年経年）

統計表・グラフ表示

統計名	家計調査 家計収支編 二人以上の世帯
表番号	010
表題	品目分類（2020年改定）（総数：金額）

▼ もっと見る

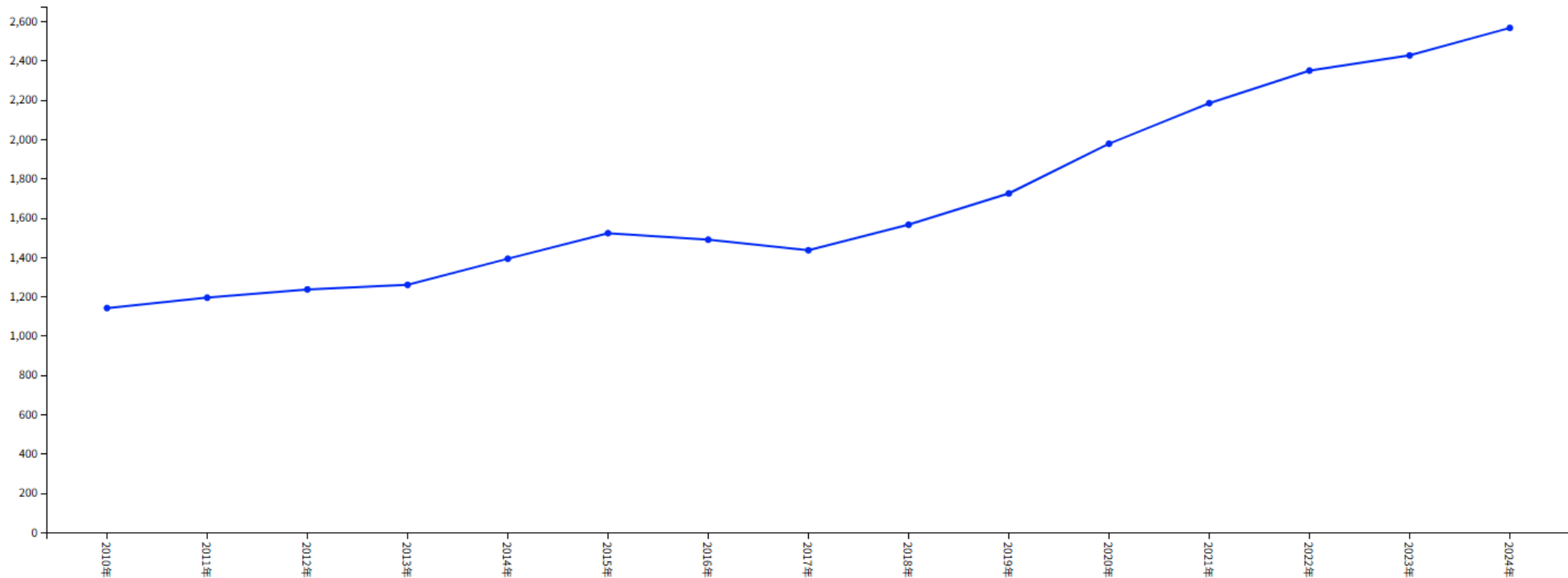
統計表表示

グラフ表示



世帯区分（年次ー二人以上の世帯）  地域区分

チョコレート菓子支出額



表示項目選択 >

レイアウト設定 >

グラフ表示設定 >

② ビジネスシーンでの活用可能性  
ー 市場分析・需要予測・地域比較などへの応用

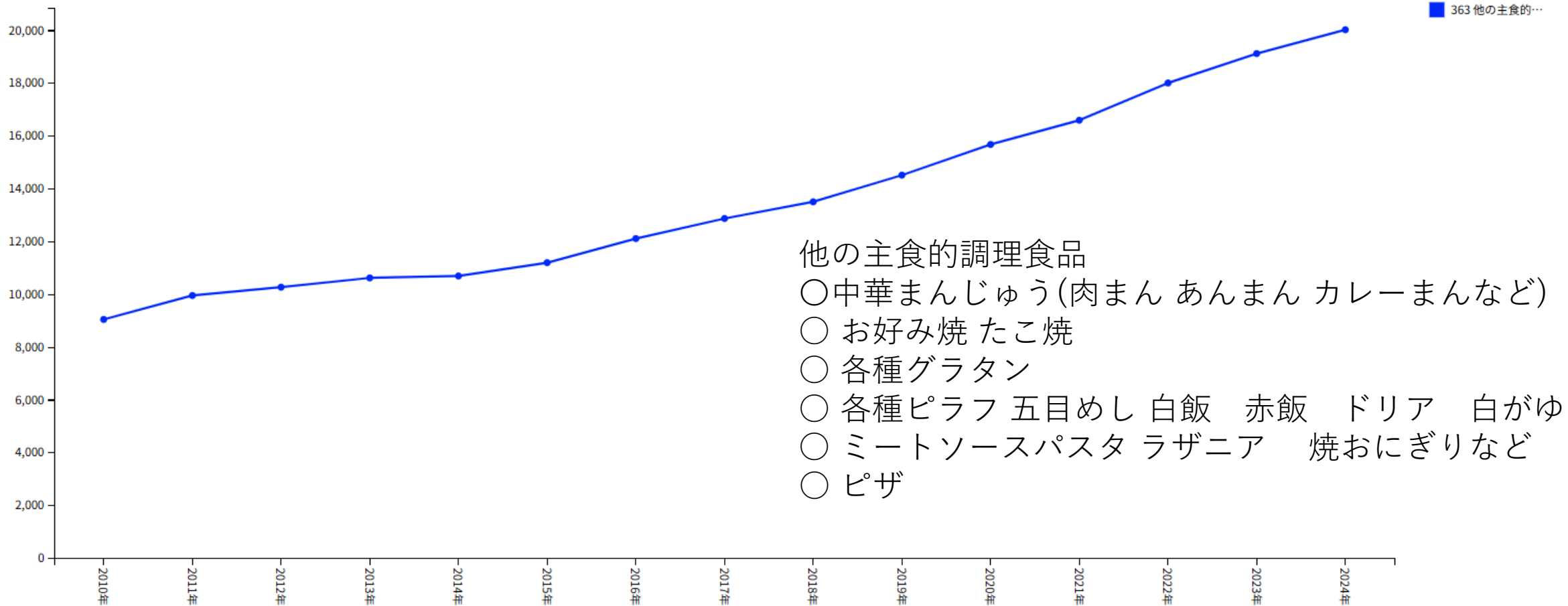
# 他の主食的調理食品支出額 (2010～2024年経年)

統計表表示

グラフ表示



表章項目 金額 世帯区分 (年次-二人以上の世帯) 二人以上の世帯 (2000年～) 地域区分 全国 再表示



表示項目選択

レイアウト設定

グラフ表示設定

② ビジネスシーンでの活用可能性  
ー 市場分析・需要予測・地域比較などへの応用

# 月謝支出額（スポーツ・音楽・語学）（2024年収入階級別）

統計表表示

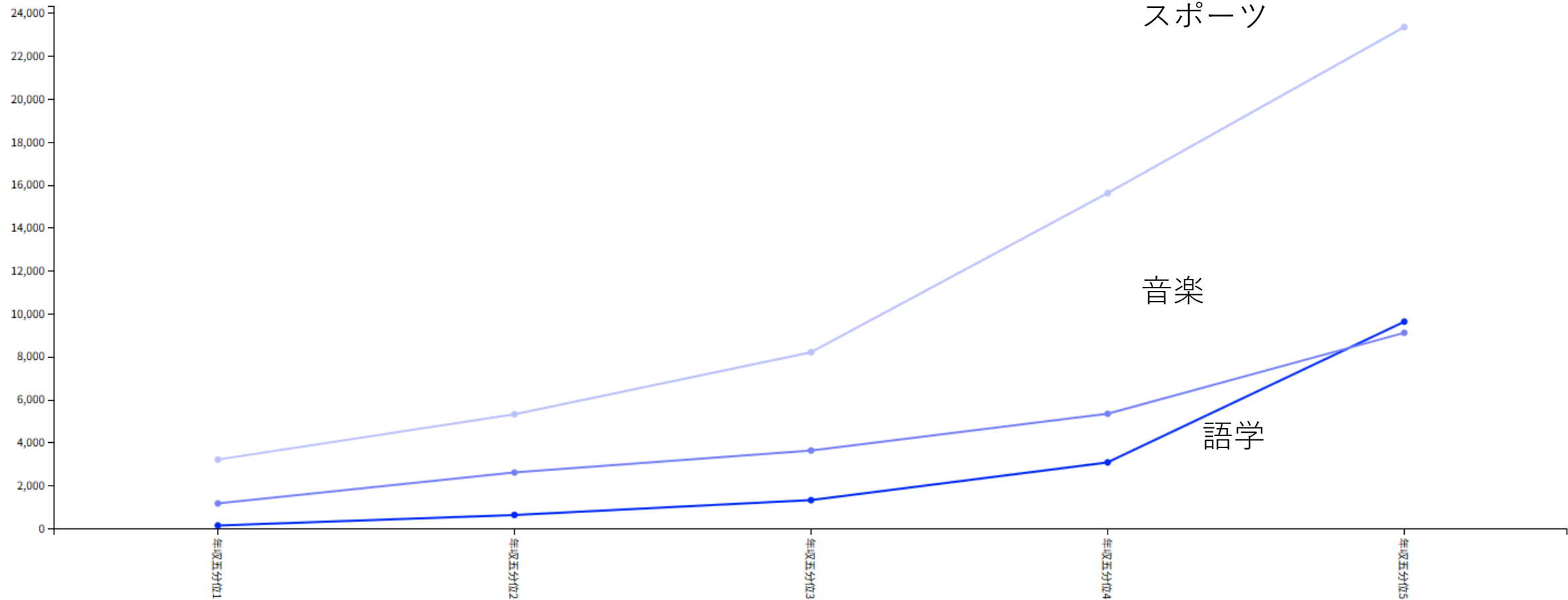
グラフ表示



表示項目 金額 地域区分 全国 時間軸 (年次) 2024年 再表示

表示項目選択  
レイアウト設定  
グラフ表示設定

月謝（スポーツ、音楽、語学）



## ② ビジネスシーンでの活用可能性 ー 市場分析・需要予測・地域比較などへの応用 酒類支出額（2024年収入階級別）

統計表表示

グラフ表示

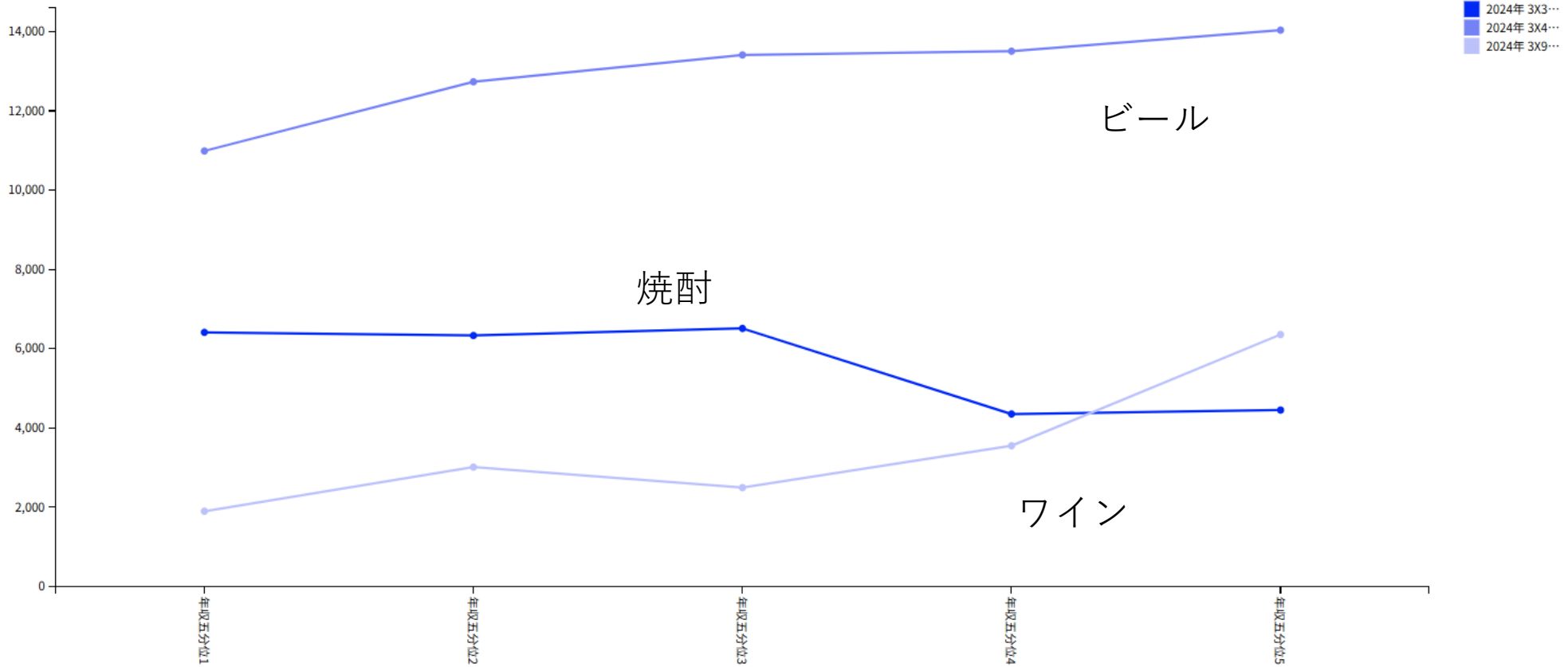
ダウンロード

表示項目選択

レイアウト設定

グラフ表示設定

表章項目 金額 世帯区分 (年次一人以上の世帯) 二人以上の世帯 (2000年～) 地域区分 全国 再表示



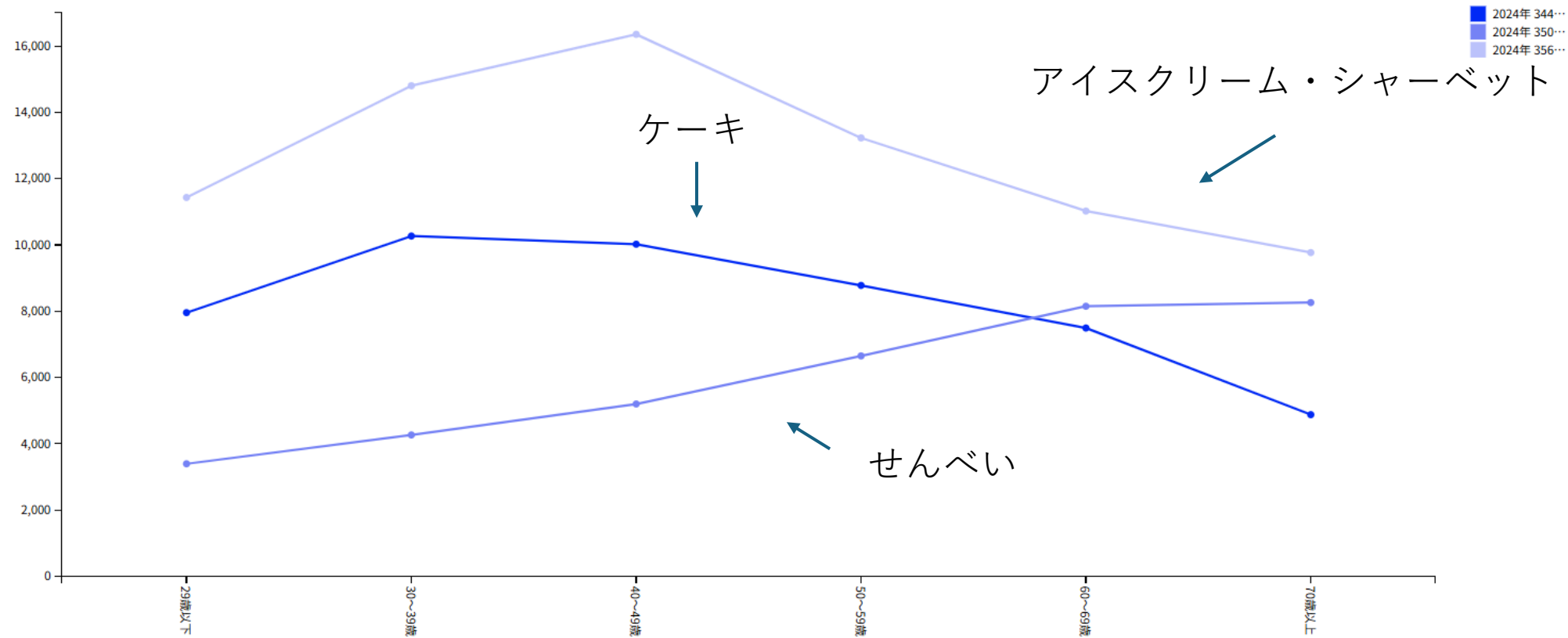
# 菓子類支出額 (2024年年齢階級別)

統計表表示

グラフ表示

ダウンロード

表示項目 金額 世帯区分 (年次一人以上の世帯) 二人以上の世帯 (2000年～) 地域区分 全国 再表示



40~49歳

# 政府統計データ活用の先行事例紹介

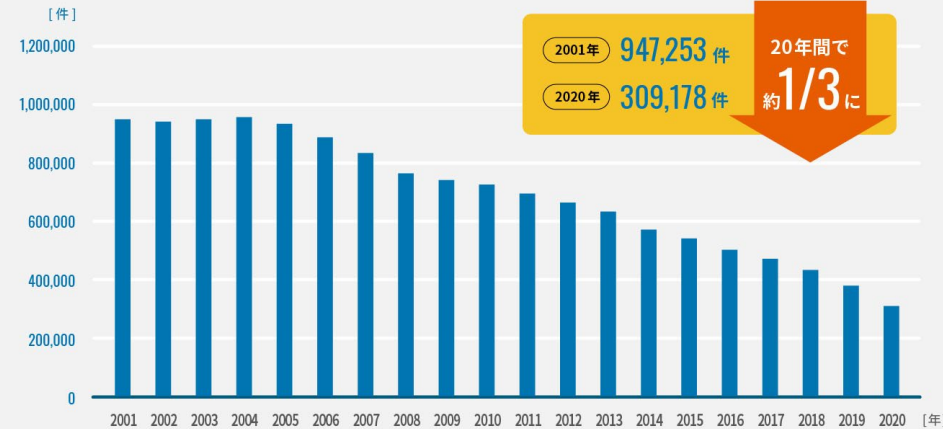
- ③ 政府統計データ活用の先行事例紹介  
 ー 民間企業での具体的導入例を紹介

## 市場動向を把握

### ナッフアクトリー（板金塗装業）

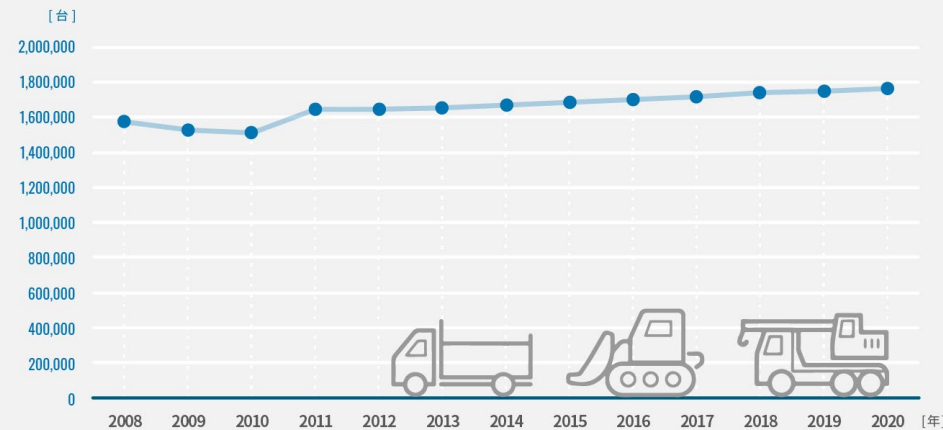
- 事故件数の減少で、板金塗装業市場の縮小を確信。
  - 自動車保有台数（国土交通省）と自動車事故（警察庁 道路の交通に関する統計）の件数
  - 安全装置付き自動車の普及状況
  - 損害保険会社の支払額の推移
- 一方で救急車や特装車など特殊車両への需要が堅調に推移。
  - 自動車保有車両数特殊用途合計（国土交通省）
- 特殊自動車の架装（装備や機器の取り付け）に業態転換。

交通事故発生状況の推移（2001年～2020年）



[資料] 道路の交通に関する統計（警察庁）

自動車保有車両数 当月末 特種(殊)用途合計数（2008年～2020年）



[資料] 自動車保有車両数（国土交通省）

（出典）経済産業省、統計利活用事例集、統計データを活用した企業人インタビュー  
[https://www.meti.go.jp/statistics/pr/rikatuyou\\_20210326/rikatuyou\\_20210326.html](https://www.meti.go.jp/statistics/pr/rikatuyou_20210326/rikatuyou_20210326.html)



③ 政府統計データ活用の先事例紹介  
ー 民間企業での具体的導入例を紹介

# 不動産取引価格を推定

## 株式会社おたに

- GEEOでは、総務省統計局や国土交通省などが提供している路線価、国勢調査、住宅・土地統計調査などの官公庁系オープンデータをベースに、独自のアルゴリズムで物件の販売価格を予想。
- 価格予想にオープンデータを取り入れることで、解析対象となるデータ量は格段に増加し、アルゴリズムの工夫次第では、今後、データから解釈できる内容を広げることができる。

(出典) デジタル庁 民間事業者による利活用事例、オープンデータ100民間事業者による事例  
[https://www.digital.go.jp/resources/data\\_case\\_study\\_private](https://www.digital.go.jp/resources/data_case_study_private)

1. なぜ今「データ分析×政府統計」か？

2. データ分析は何から始めればいいの？

3. Excelを使った政府統計データの分析(演習)

地図上の対象地点の不動産価格を表示（売買・投資物件、自宅の資産価値の検討等に有効）



時系列で価格表示も可能。

- ③ 政府統計データ活用の先行事例紹介  
 ー 民間企業での具体的導入例を紹介

# 揺れやすさ指標の作成

## 大和ハウス工業株式会社

- 建設業者や住宅購入者向けに提供される地震リスク評価サービス。
- 所在地を入力すると、J-SHIS の地震動予測値や地盤データを基に揺れやすさを算出する。

(出典) デジタル庁 民間事業者による利活用事例、オープンデータ100民間事業者による事例  
[https://www.digital.go.jp/resources/data\\_case\\_study\\_private](https://www.digital.go.jp/resources/data_case_study_private)

→国勢調査データを組み合わせると、人口密度を確認したうえで災害時の危険度を評価することも可能。

1. なぜ今「データ分析 × 政府統計」か？

2. データ分析は何から始めればいいの？

3. Excelを使った政府統計データの分析(演習)

大阪・梅田に住宅を購入する予定の場合

The screenshot shows a three-step process for earthquake risk evaluation. Step 1, '住所を入力してください' (Please enter your address), has dropdown menus for '都道府県' (大阪府), '市町村' (大阪市北区), '町丁目' (梅田三丁目), and '番地' (3-5). Step 2, '解析が終了しました' (Analysis completed), asks for the analysis period, with '20年間' selected. Step 3, '評価結果' (Evaluation Results), shows a map and a table of results:

震度	発生確率
震度5弱以上	91.6%
震度5強以上	71.7%
震度6弱以上	38.9%
震度6強以上	8.7%

建設予定地の住所を入力することで、地震発生確率や予測震度を診断してくれる

## ココゆれでこう変わった！

- 地震の発生確率や予測震度が簡単に読み取ることができるになり、顧客は専門的な知識を持たなくても地震リスクを自身で調べられるようになった



## 2. データ分析は何から始めればいいのか？

### ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎
- 相関分析：2つの指標の関係を捉える
- 回帰分析：2つの変数の関係を捉える

### ②データ分析の基本姿勢

- 思い込みに左右されない判断
- グラフの目的に応じた使い方を学ぶ

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 見え方の基礎

- 平均値、中央値、最頻値
- 標準偏差
- 相関

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 多くのデータを代表値で表す

- 平均値：すべてのデータを足して、データの数で割る
- 中央値（メジアン）：データを値の大きさの順に並べた場合の真ん中の値
- 最頻値（モード）：データセットの中で最も頻繁に出現したデータ

それぞれ、エクセルで計算できる。

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 平均値の計算（その1）

- 平均値を求めてください。
- 平均値は全部足して、個数で割ったもの。
- $55 \div 11 = 5$

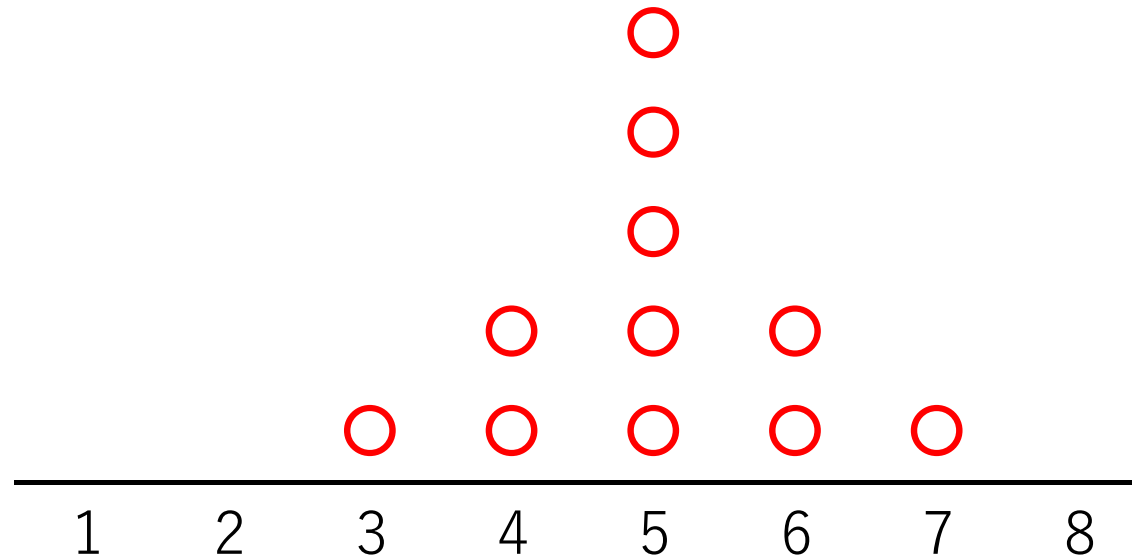
5 7 5 6 3 4 4 5 6 5 5

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 中央値と最頻値の計算 (その1)

- 中央値：11サンプルなので、上から6番目→5
- 最頻値：5が5回あるので、5



## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 平均値の計算（その2）

- 平均値を求めてください。
- 平均値は全部足して、個数で割ったもの。
- $55 \div 11 = 5$

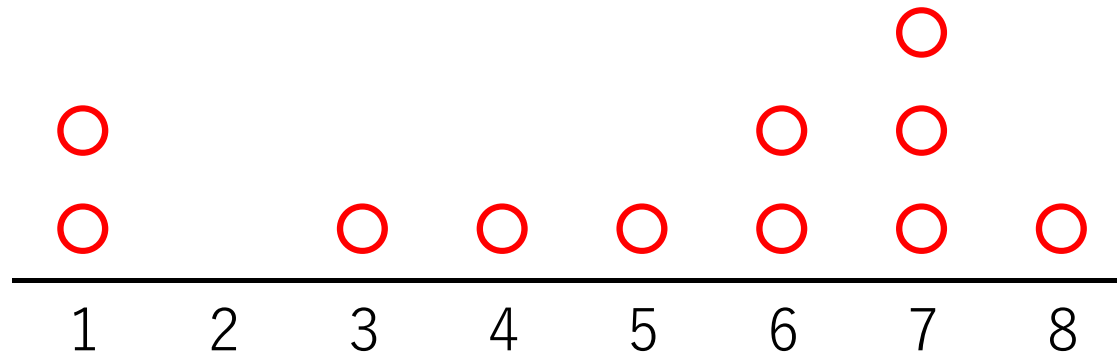
7 7 1 6 3 5 1 7 8 4 6

①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 中央値と最頻値の計算 (その2)

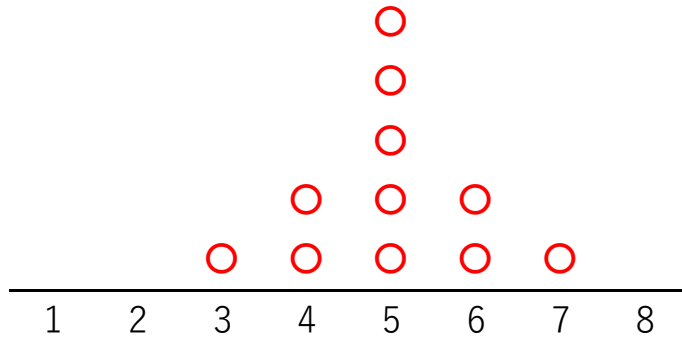
- 中央値：11サンプルなので、上から6番目→6
- 最頻値：7が3回あるので、7



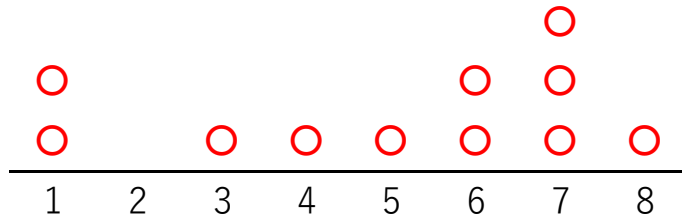
## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 分布が変わると、代表値は変化する



平均値：5  
中央値：5  
最頻値：5

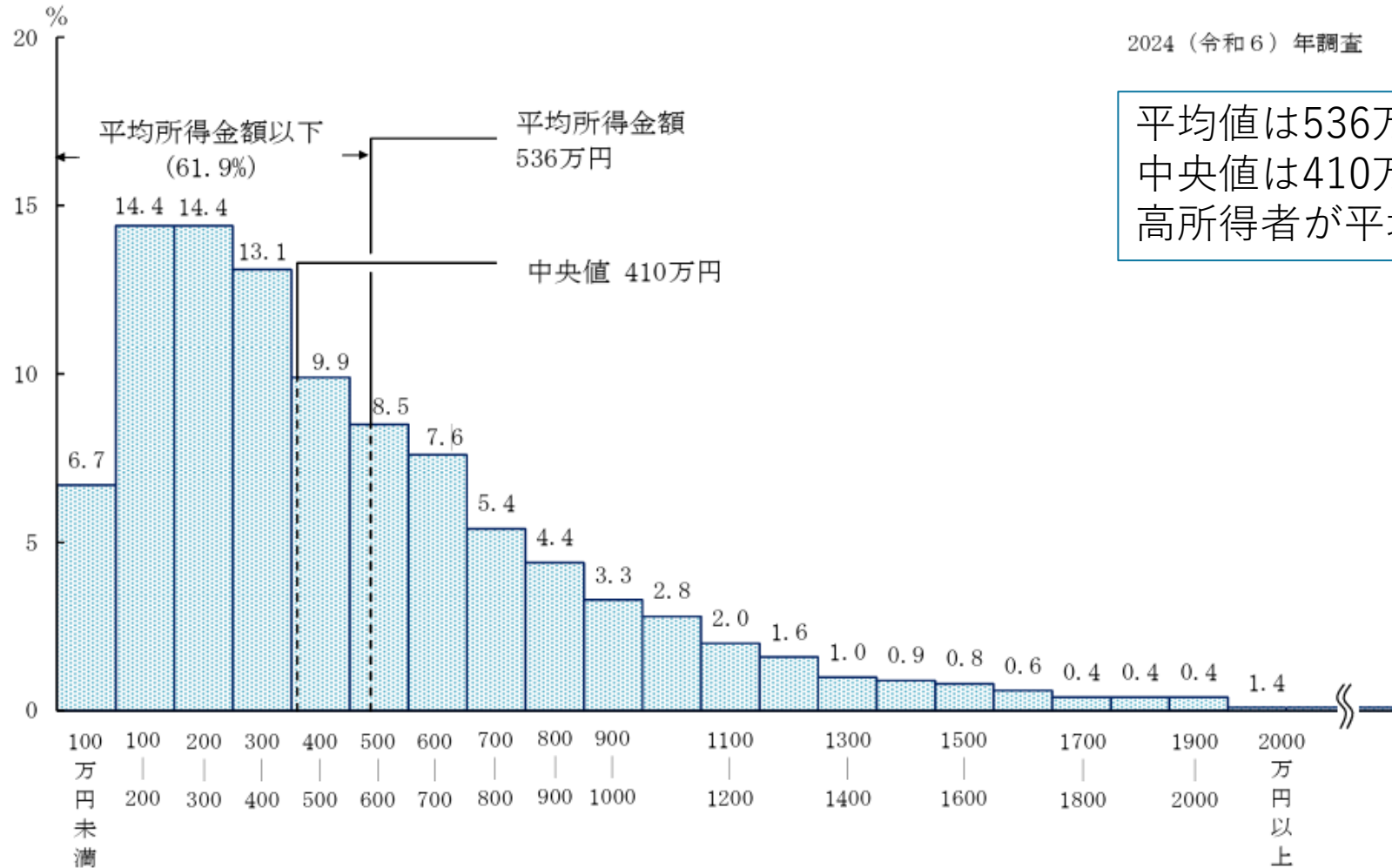


平均値：5  
中央値：6  
最頻値：7

# ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

図9 所得金額階級別世帯数の相対度数分布



平均値は536万円だが、中央値は410万円。高所得者が平均を引き上げている。

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

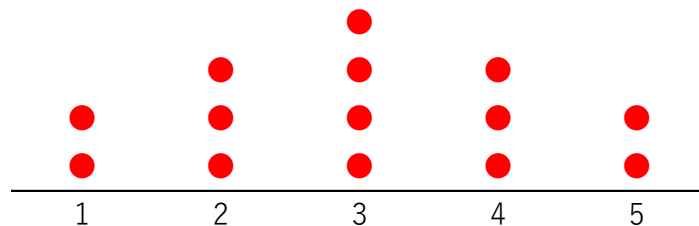
# 「ばらつき」を考える理由

- 代表値（平均・中央値・最頻値）に加えて、データの「ばらつき」を知りたいことは多い。

例1：アンケート調査（5が満足度高）

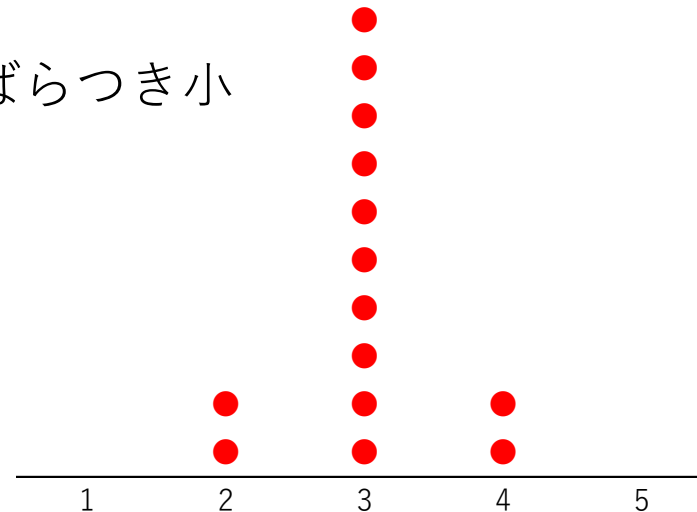
平均はどちらも3だが…

ばらつき大



満足度に大きな差

ばらつき小



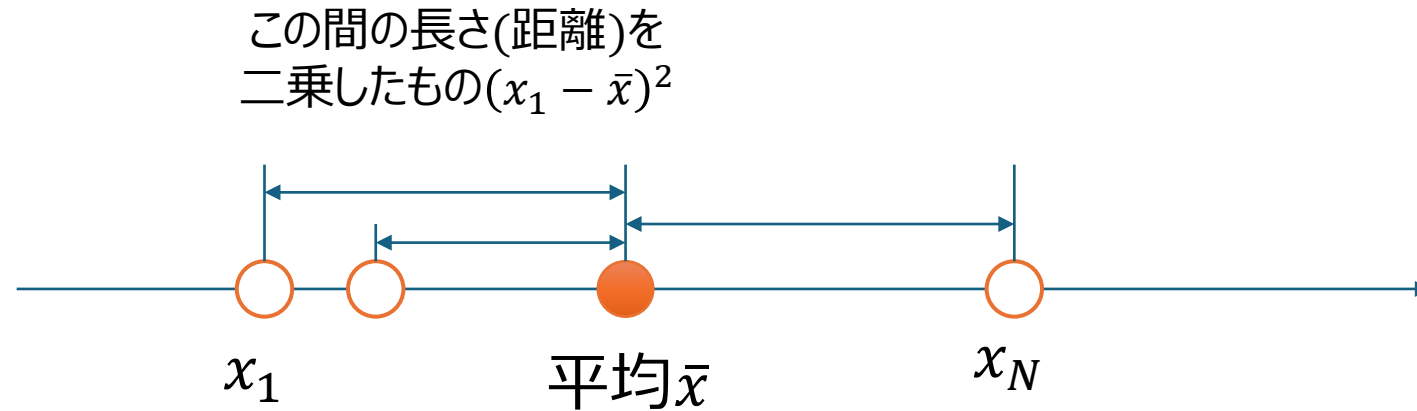
満足度はだいたい同じ

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# ばらつき（分散）を測る方法

- 全データが平均的に「平均 $\bar{x}$ とどれくらい離れているか？」を測る。



- 分散が大きい→平均値と大きく違う数が多い→広がっている。
- 差を単純に足すと、プラスとマイナスが打ち消してゼロになるので、それぞれ二乗する。

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 分散の公式

- データ  $x_1, x_2, \dots, x_N$  の分散
  - 「平均値との差の二乗」の平均

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

平均値

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 標準偏差もばらつき具合を表す

## • 例

- 分散が100なら、標準偏差は10
- 分散が2なら、標準偏差は $\sqrt{2}$
- 分散が0なら、標準偏差も0

## • なので、標準偏差を二乗したら分散

- 標準偏差が10なら、分散は100
- 標準偏差が $\sqrt{2}$ なら、分散は2

• ばらつき度合いを表すという意味では分散と同じ。分散の平方根をとったもの。

• あるデータがどのくらい平均から離れているかを表すときの物差しとなる。

• ばらつきの1単位。

(例) 標準偏差の0.5倍ならあまり平均からはなれていない。2倍なら結構離れている…など。

①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

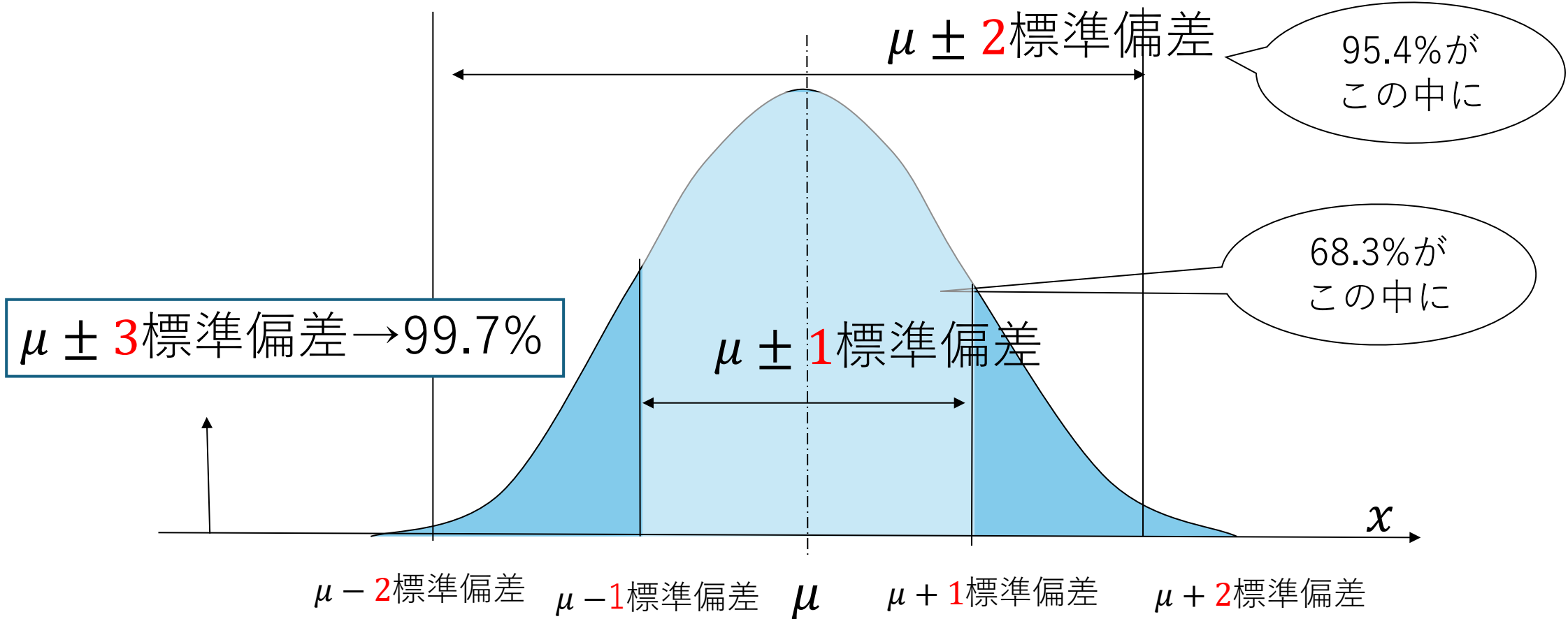
# 標準偏差の使い方

1. なぜ今「データ分析 × 政府統計」か？

2. データ分析は何から始めればいいの？

3. Excelを使った政府統計データの分析(演習)

<ポイント>	
1標準偏差	68.3%(7割)
2標準偏差	95.4%
3標準偏差	99.7%



プラスマイナス1標準偏差なら7割、プラスマイナス2標準偏差ならほぼ全部。

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 標準偏差

- (出典) 厚生労働省「2023年 国民健康・栄養調査」40歳～49歳

	性別	平均値	標準偏差	平均-2 標準偏差	平均+2 標準偏差
体重 (kg)	男性	70.4	12.1	46.2	94.6
	女性	55.8	10.3	35.2	76.4
歩数 (歩)	男性	8,039	4,871	-1,703	17,781
	女性	7,172	4,218	-1,264	15,608

①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

# 歩数（男性：40歳～49歳）

1. なぜ今「データ分析 × 政府統計」か？

2. データ分析は何から始めればいいの？

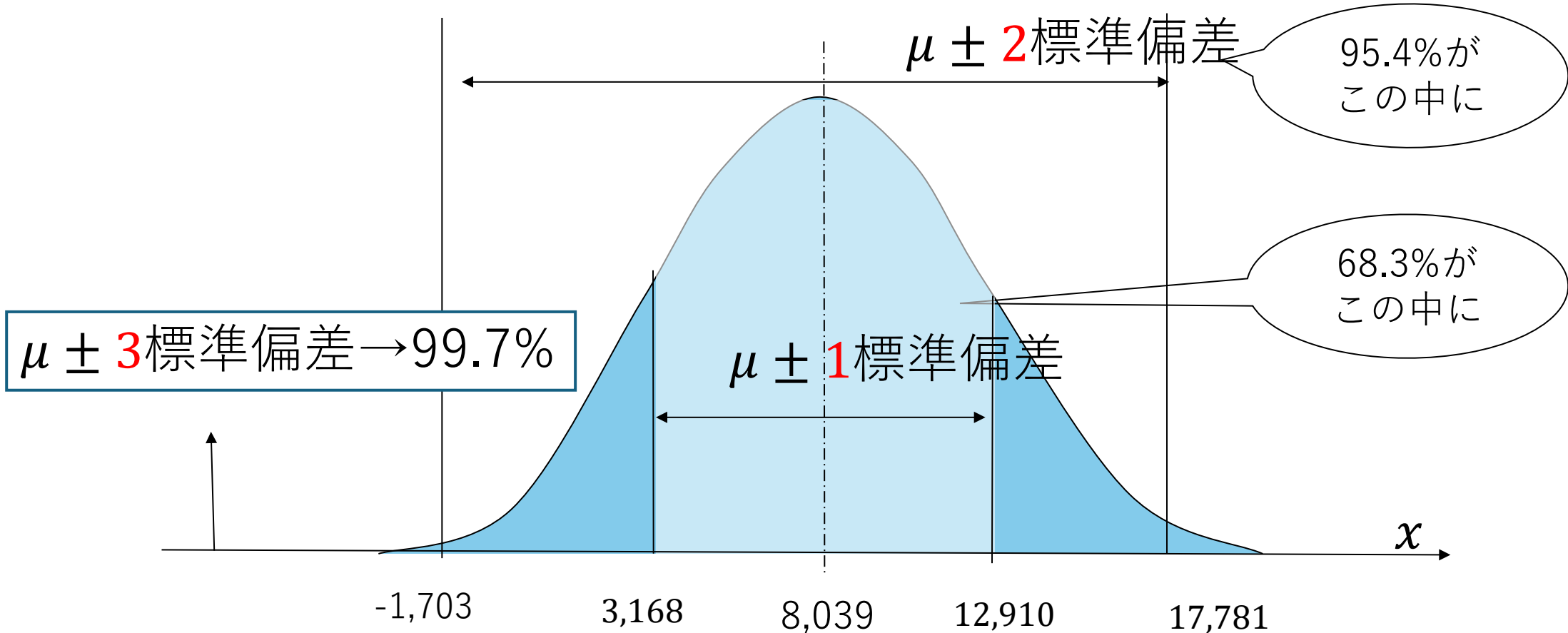
3. Excelを使った政府統計データの分析(演習)

<ポイント>

1標準偏差 68.3%(7割)

2標準偏差 95.4%

3標準偏差 99.7%



プラスマイナス1標準偏差なら7割、プラスマイナス2標準偏差ならほぼ全部。

## ①データの「見え方」を知る

- 平均値・中央値・最頻値・標準偏差などの基礎

## エクセルの関数を使って計算できます

代表値	関数
平均値	=AVERAGE(セルの範囲)
中央値	=MEDIAN(セルの範囲)
最頻値	=MODE(セルの範囲)
分散	=VAR.P(セルの範囲) =VARP(セルの範囲)
標準偏差	=STDEV.P(セルの範囲) =STDEVP(セルの範囲)

	A	B	C
1	名前	得点	
2	Aさん	67	
3	Bさん	88	
4	Cさん	82	
5	Dさん	71	
6	平均	77	
7			
8			
9			
10			

=AVERAGE(B2:B5)

セルの範囲の示し方→

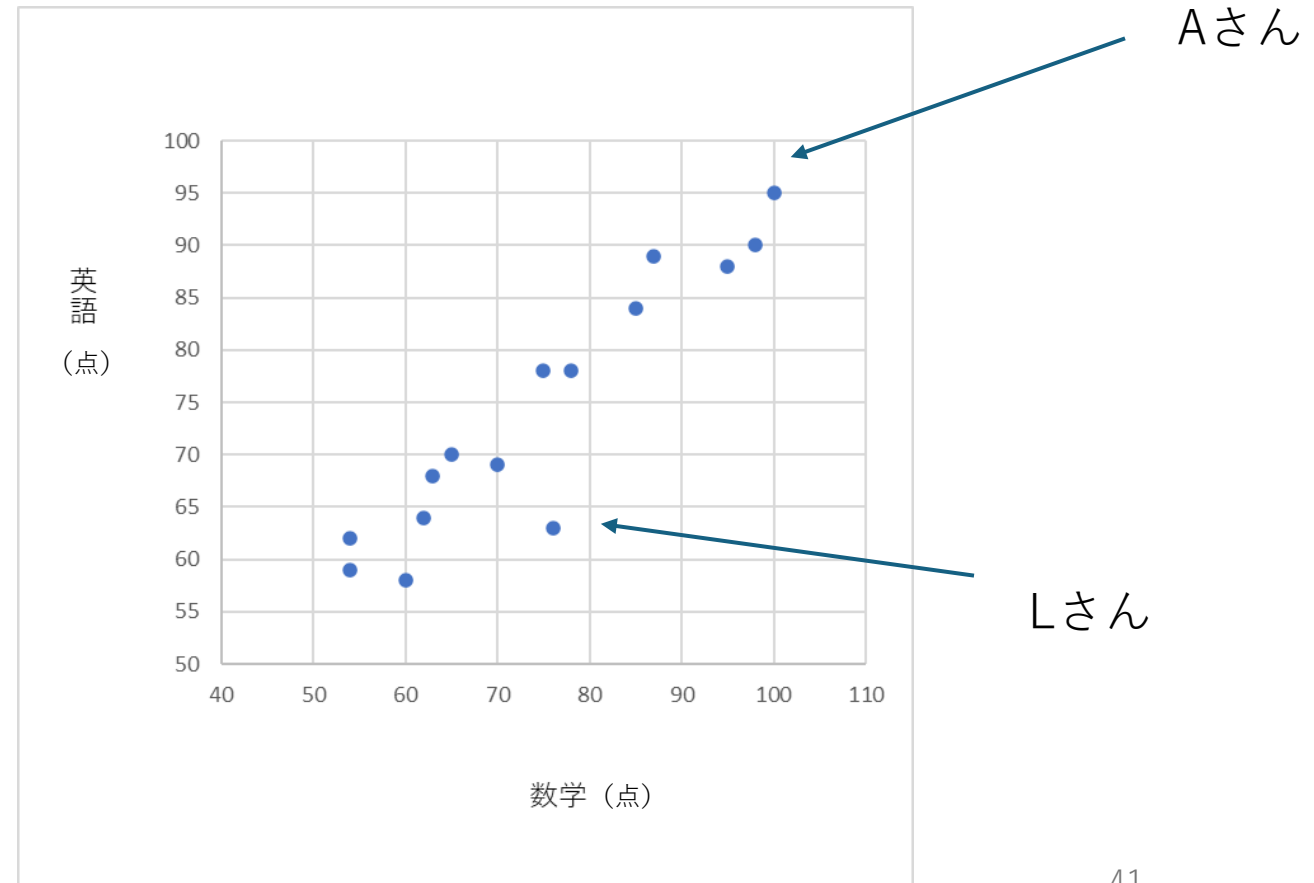
## ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

# 散布図

- 2つの変数の関係を表す。1つの変数が増えた時、もう1つの変数がどうなるかがわかる。

	数学 (点)	英語 (点)
A	100	95
B	98	90
C	95	88
D	87	89
E	85	84
F	78	78
G	75	78
H	65	70
I	70	69
J	63	68
K	62	64
L	76	63
M	54	62
N	54	59
O	60	58



# ①データの「見え方」を知る

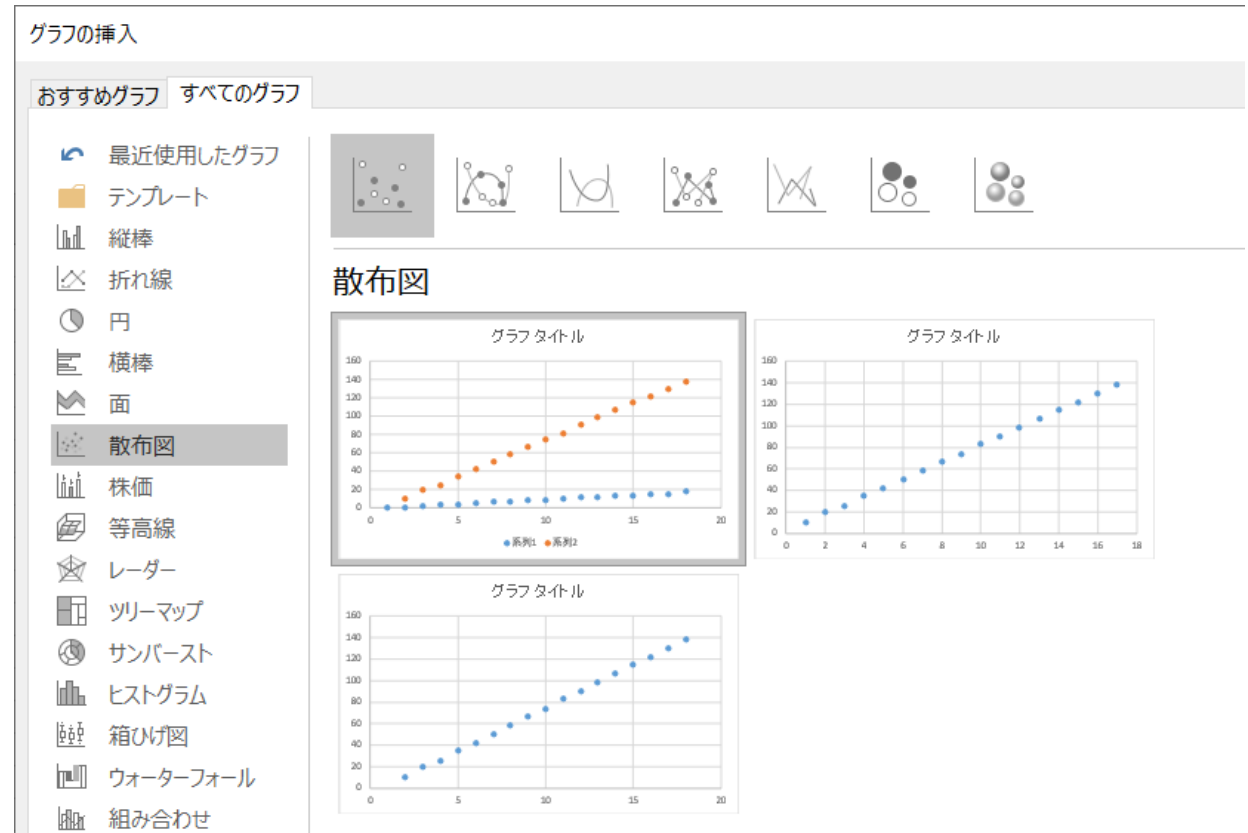
- 相関分析：2つの指標の関係を捉える

# 散布図の作り方

• Microsoft Excelのグラフ機能に標準で搭載されている。



データの範囲を選択→挿入タブ→グラフ挿入画面で散布図を選択。



## ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

# 「相関」とは

## 2つの量に一定の関係があること

- 2つの量の関係性
  - 経済分析では因果関係を知りたいことが多い
  - 因果関係とまではいかななくても、関係性があるかどうかをまず調べる
- 例
  - 気温とアイスの売り上げ高の関係性
  - 広告費と売上高
  - 価格と需要量
  - 運動量と体脂肪率
  - 降水量と傘の売上高

# 3種類の相関

- 正の相関

- $x_1 \rightarrow$ 大、 $x_2 \rightarrow$ 大
- 例：身長と体重

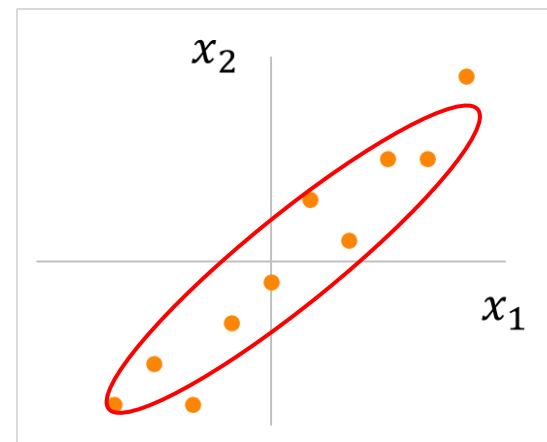
- 負の相関

- $x_1 \rightarrow$ 大、 $x_2 \rightarrow$ 小
- 例：身長と100m走のタイム

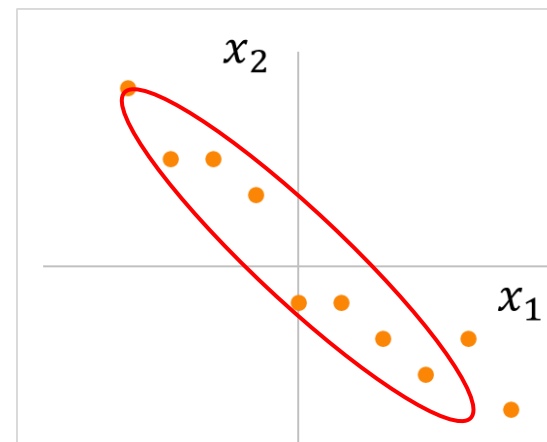
- 無相関

- $x_1 \rightarrow$ 大、 $x_2 \rightarrow$ 特段の傾向無し
- 要するに、 $x_1$ と $x_2$ は無関係
- 例：身長と年収

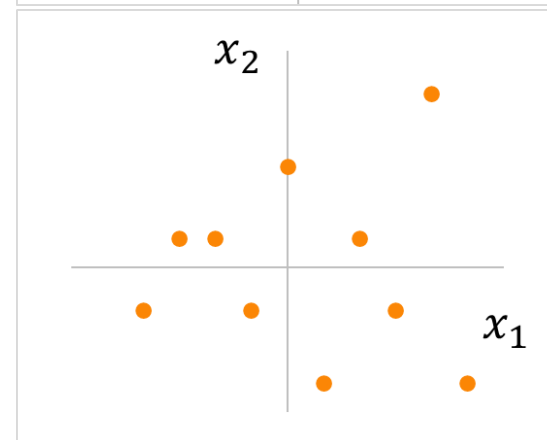
正の相関



負の相関



無相関



# 相関係数

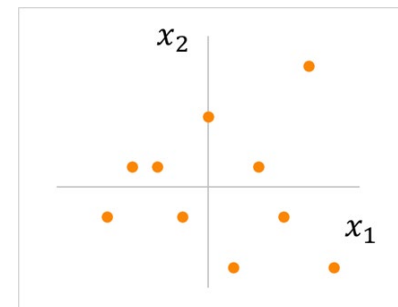
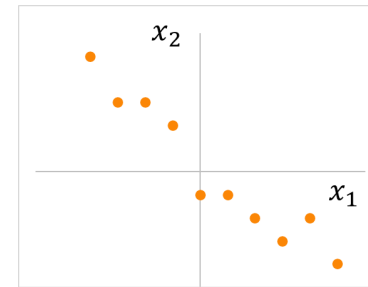
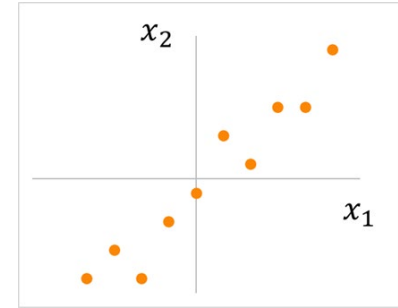
分子のプラスマイナスに注目



$$\rho = \frac{(x_1 \times x_2) \text{の平均値}}{\sqrt{x_1 \text{の分散} \cdot x_2 \text{の分散}}}$$

簡単に説明するため、 $x_1$ と $x_2$ の平均値がゼロの場合

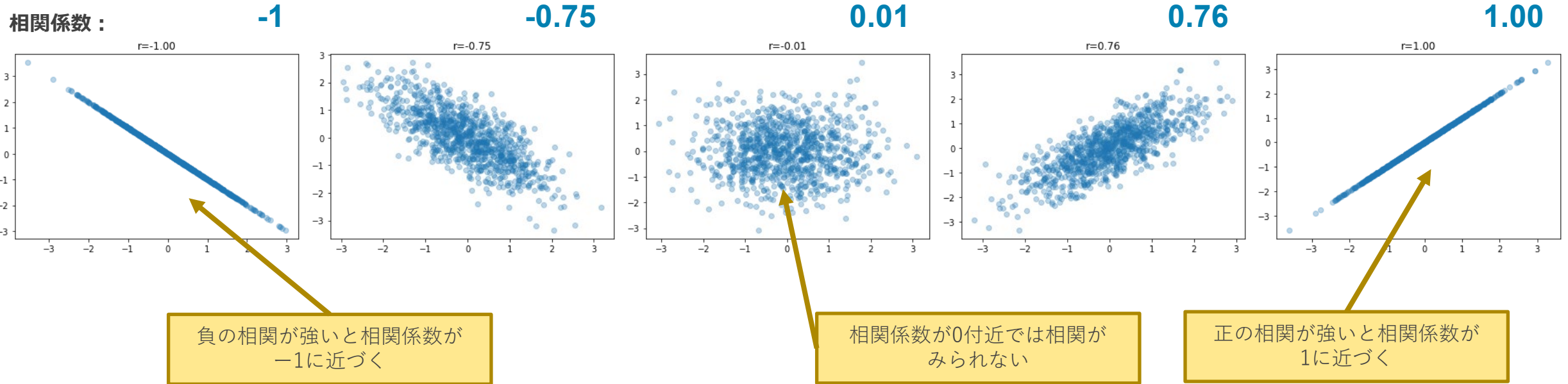
- 正の相関： $x_1$ と $x_2$ が同じ符号(+と+、-と-)になりやすい  
→ $x_1 \times x_2$ は正になりやすい
- 負の相関： $x_1$ と $x_2$ が逆の符号になりやすい  
→ $x_1 \times x_2$ は負になりやすい
- 無相関： $x_1$ と $x_2$ の符号は同じだったり逆だったり  
→ $x_1 \times x_2$ も正だったり負だったり



①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

# 相関係数の大きさとばらつき具合の様子



## ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

# 相関を確認する

- 相関係数の関数はCORREL
- 2つの変数のセルを選ぶ
- =CORREL(セルの範囲,セルの範囲)

The image shows an Excel spreadsheet with two columns of data. The first column (C6:C10) contains the values 1, 2, 3, 4, and 5. The second column (D6:D10) contains the values 2, 4, 7, 9, and 11. A formula bar at the top right shows the formula =CORREL(C6:C10,D6:D10). The cells C6:C10 are highlighted with a blue border, and the cells D6:D10 are highlighted with a red border.

1	2
2	4
3	7
4	9
5	11

## ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

# コーヒーチェーン店出店数と総人口の相関係数

## データの収集

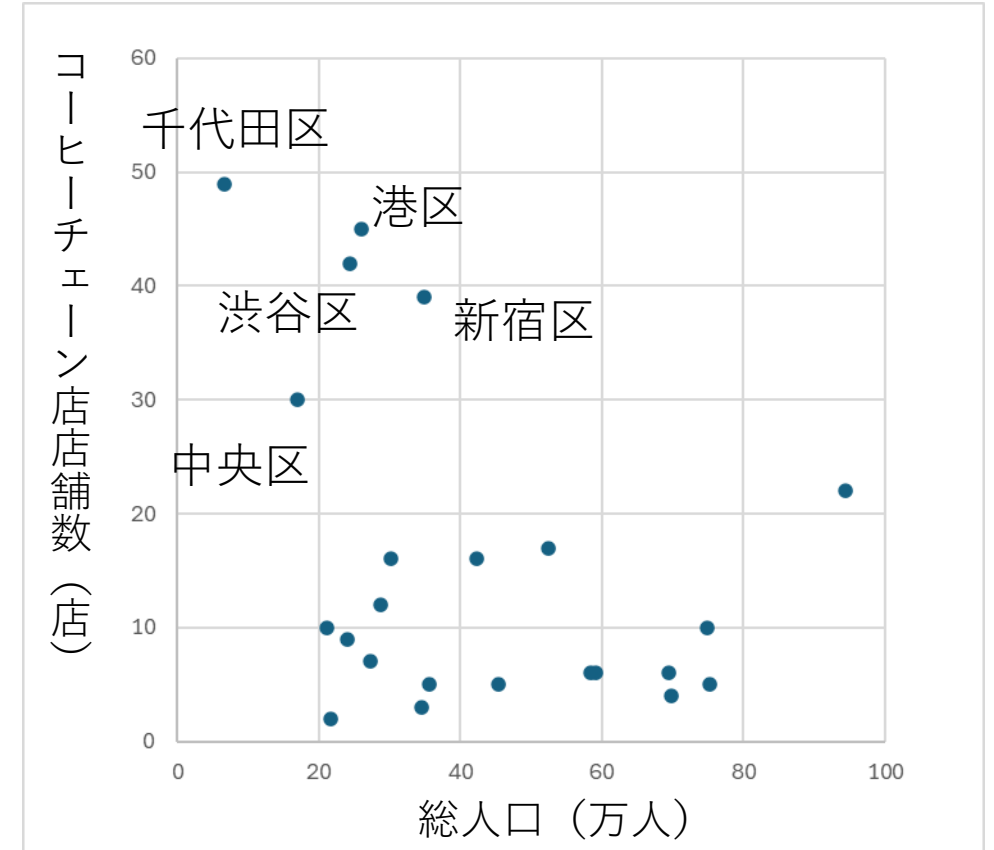
- コーヒーチェーン店出店数
  - ホームページの店舗一覧から入力
- 総人口（e-Statを利用）
  - 統計データを活用する・地域
  - 市区町村データ・データ表示
  - 東京都・特別区
  - 基礎データ・A人口・世帯
  - A1101 総人口

### ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

## コーヒーチェーン店出店数と総人口の相関係数

- 散布図を描く
- 相関係数 =CORREL( )
- = -0.40



## ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

# コーヒーチェーン店出店数と昼間人口の相関係数

## データの収集

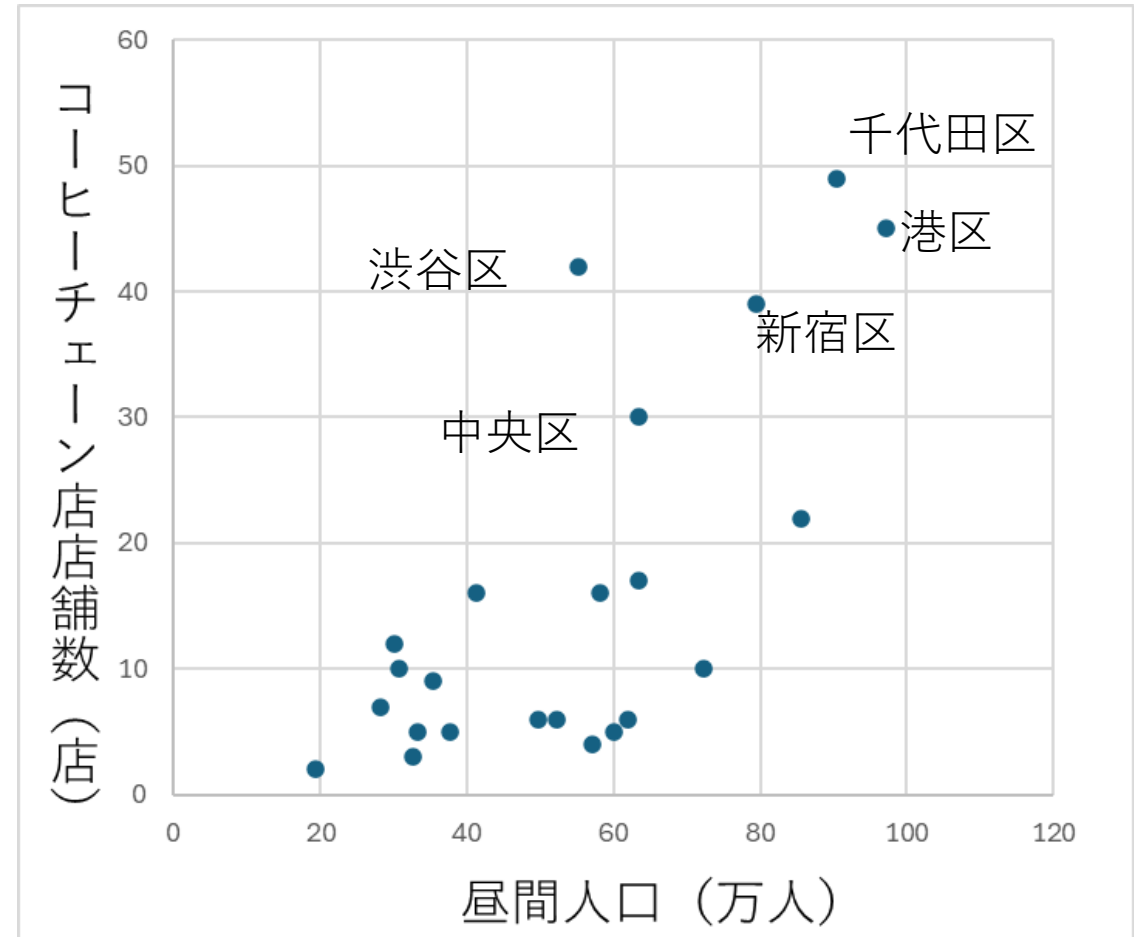
- コーヒーチェーン店出店数
  - ホームページの店舗一覧から入力
- 昼間人口（e-Statを利用）
  - 統計データを活用する・地域
  - 市区町村データ・データ表示
  - 東京都・特別区
  - 基礎データ・A人口・世帯
  - A6107 昼間人口

### ①データの「見え方」を知る

- 相関分析：2つの指標の関係を捉える

## コーヒーチェーン店出店数と昼間人口の相関係数

- 散布図を描く
- 相関係数 =CORREL( )
- =0.71



## ①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 回帰分析とは

- データ同士を結び付ける「魔法」

## ①原因の解明

- ある変数がほかの変数の原因になっているかどうかを判断することができる。

## ②予測

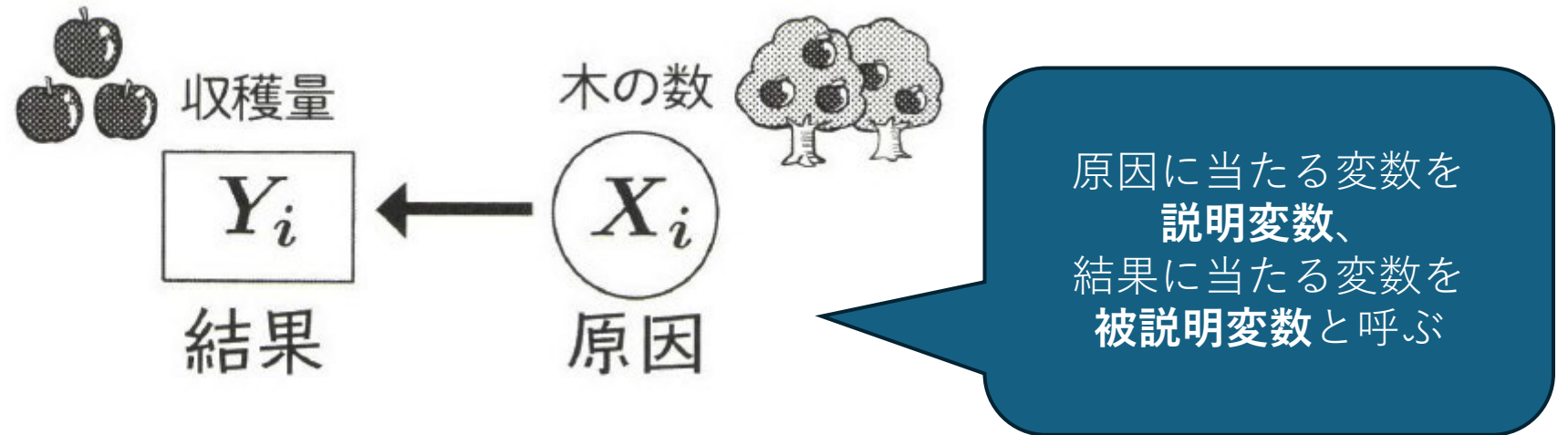
- ある変数を使って、ほかの変数を予測することができる。

①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 回帰分析は、2つの変数の関係を式で表す

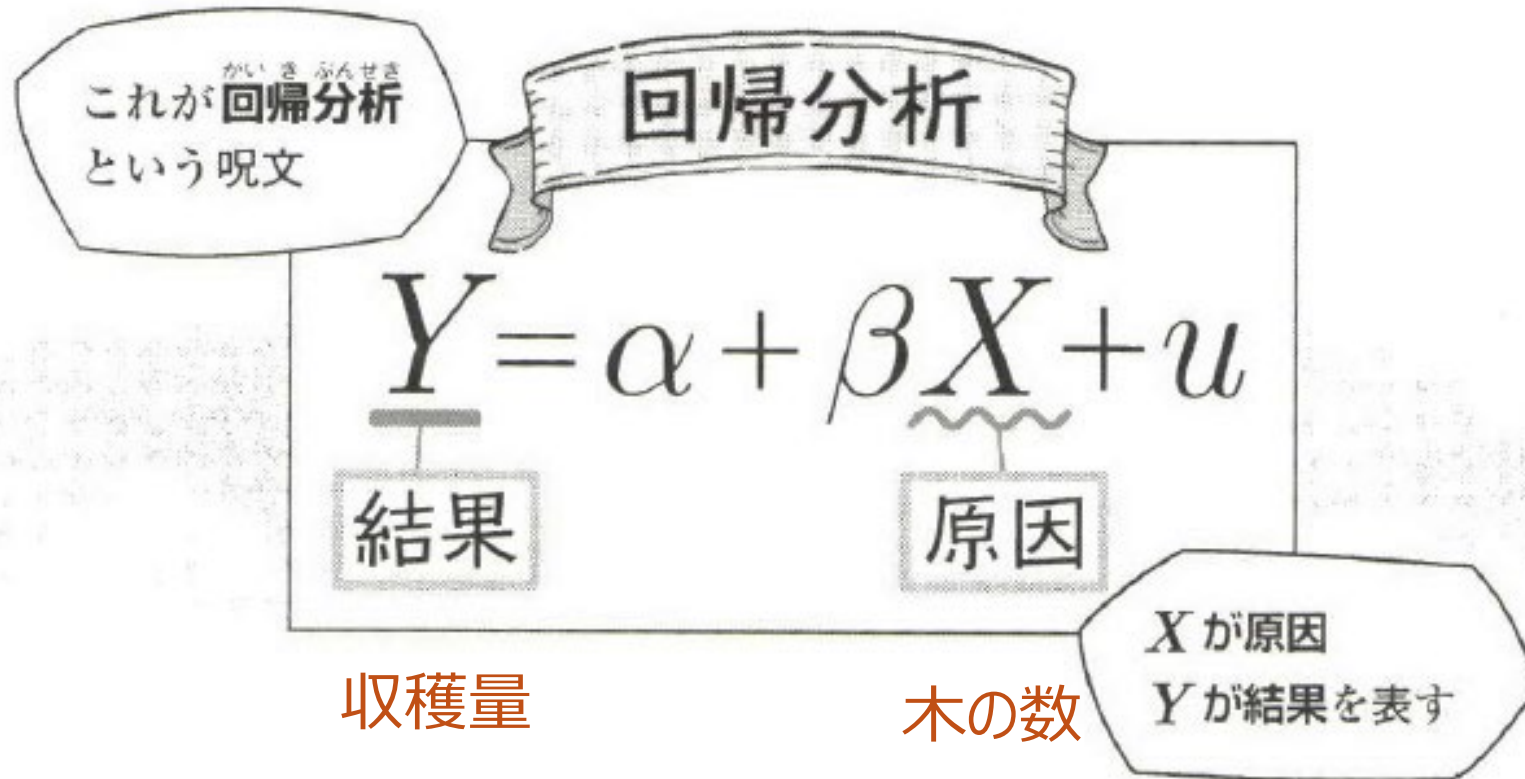
木の数が増えたら、収穫量はどうなるだろうか？



①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 2つの変数を式で関係づけるのが回帰分析



木の本数が〇〇本の時、  
収穫量は何個になるか？  
を予測することができる

①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 係数と誤差項

$$Y_i = \alpha + \beta X_i + u_i$$

係数

誤差項  
(かく乱項)

回帰式を完成するためには、 $\alpha$ や $\beta$ を知る必要がある。  
 $X$ から計算した答えはぴったり $Y$ にならないため、誤差項も必要。

①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 単回帰分析

$$Y_i = \alpha + \beta X_i + u_i$$



タンカイキ！

$$Y_i = 99 + 0.52X_i$$

$\alpha$ の推定値

$\hat{\alpha}$

$\beta$ の推定値

$\hat{\beta}$

$\alpha$ や $\beta$ は「最小二乗法」  
という方法で計算できる。

## ①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 具体的な例で考える

- 大学生の毎月のアルバイト収入と毎月の支出金額の例

	X	Y
	アルバイト収入 (万円)	支出額 (万円)
Aさん	4.0	3.1
Bさん	8.0	4.8
Cさん	12.0	6.9
Dさん	2.0	2.0
Eさん	6.0	3.9

回帰分析を使うと…

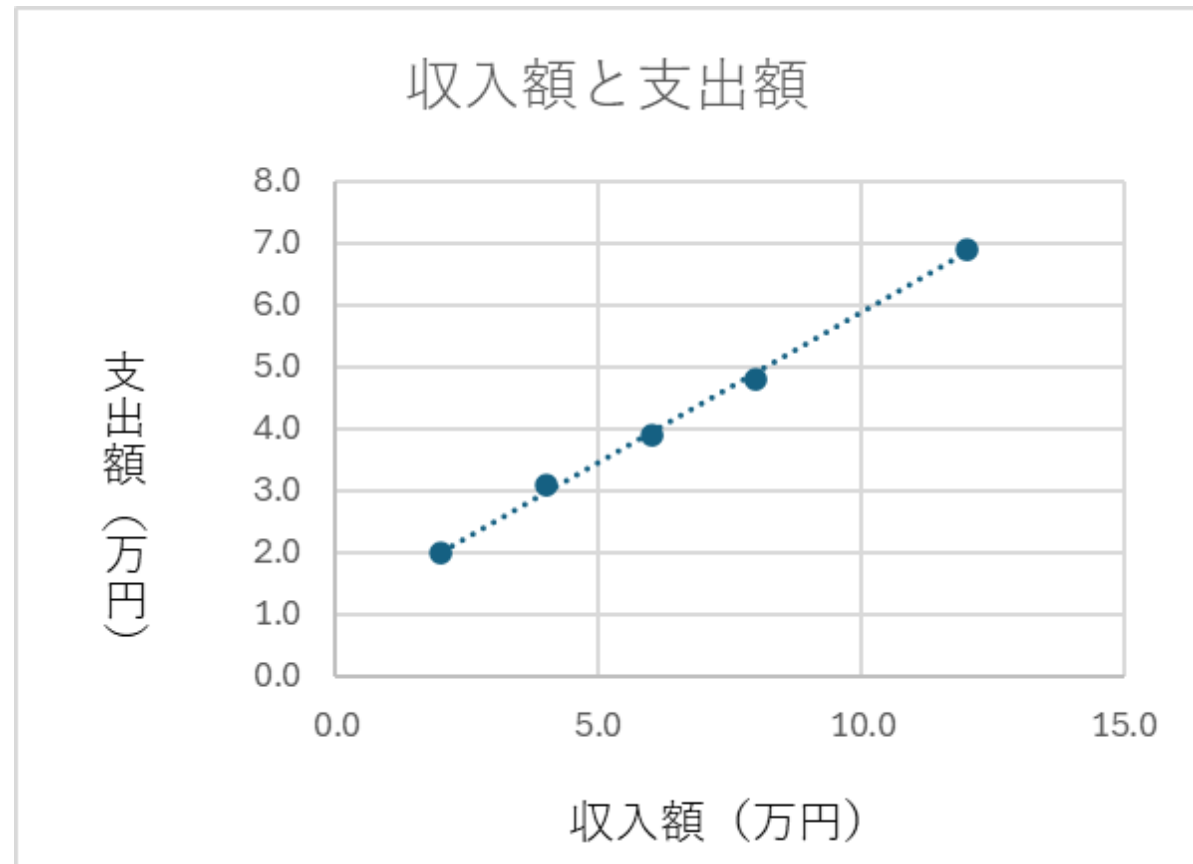
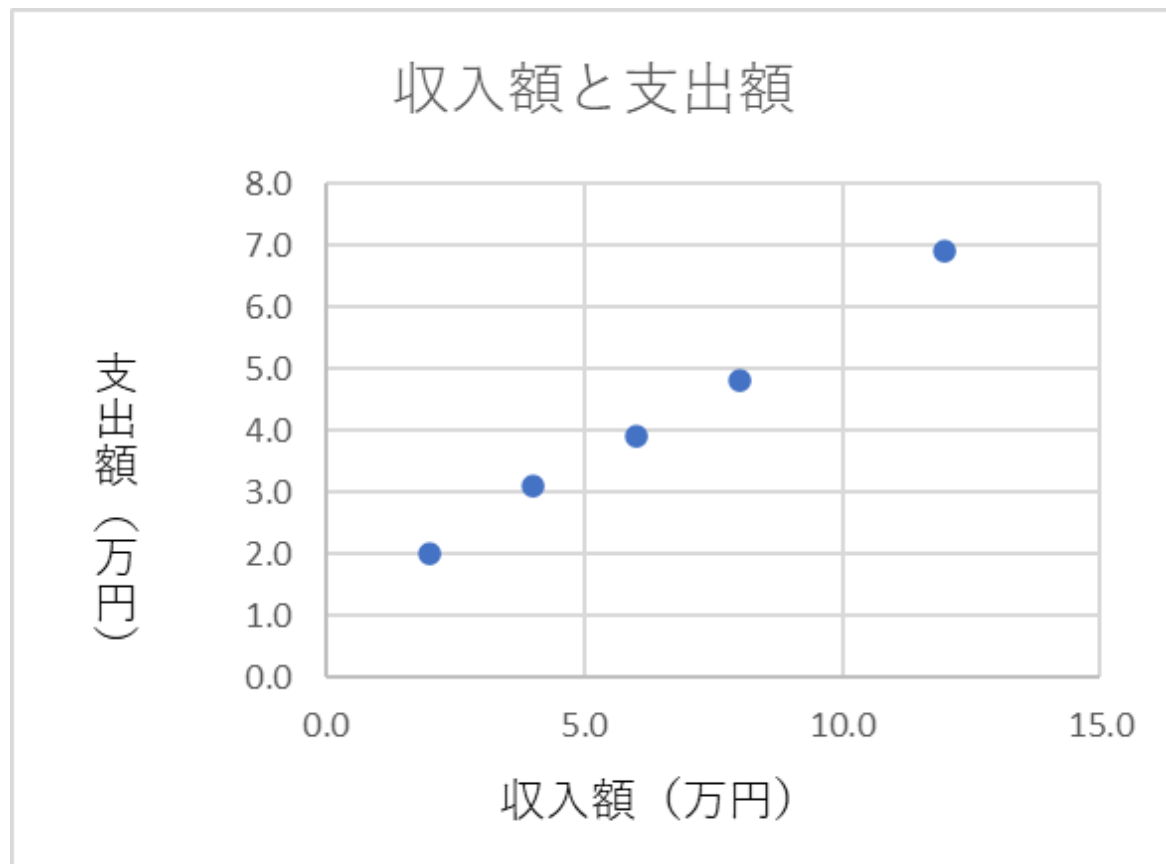
- 「収入が多い人は支出が多い。」といことを確認できる。
- 5万円の収入の人はどのくらい支出するかを予測することができる。

## ①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 散布図で描くと

直線が引けそうなので、  
1次関数の式で表せそう。



- ・ 最小二乗法を使えば、2つの関係を図示できる。
- ・ また、切片と傾きが推定できて予測ができる。

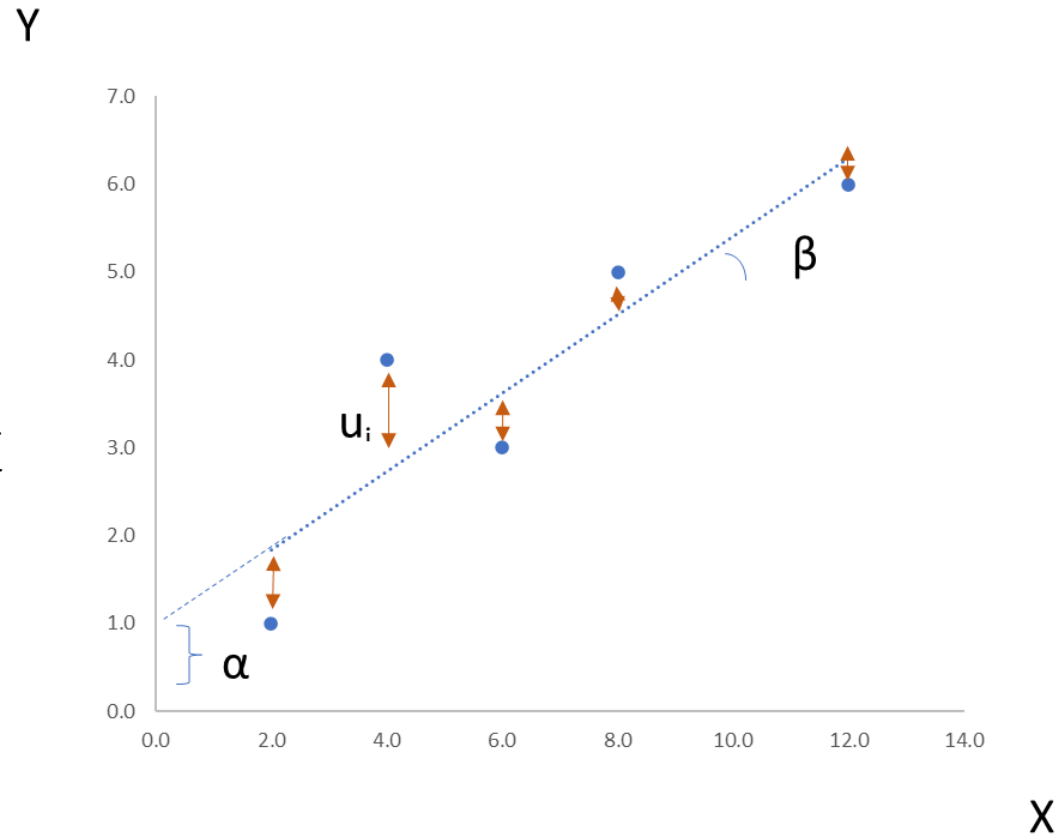
## ①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

## 実際のデータに当てはめて推定値を求める

$$Y_i = \alpha + \beta X_i + u_i$$

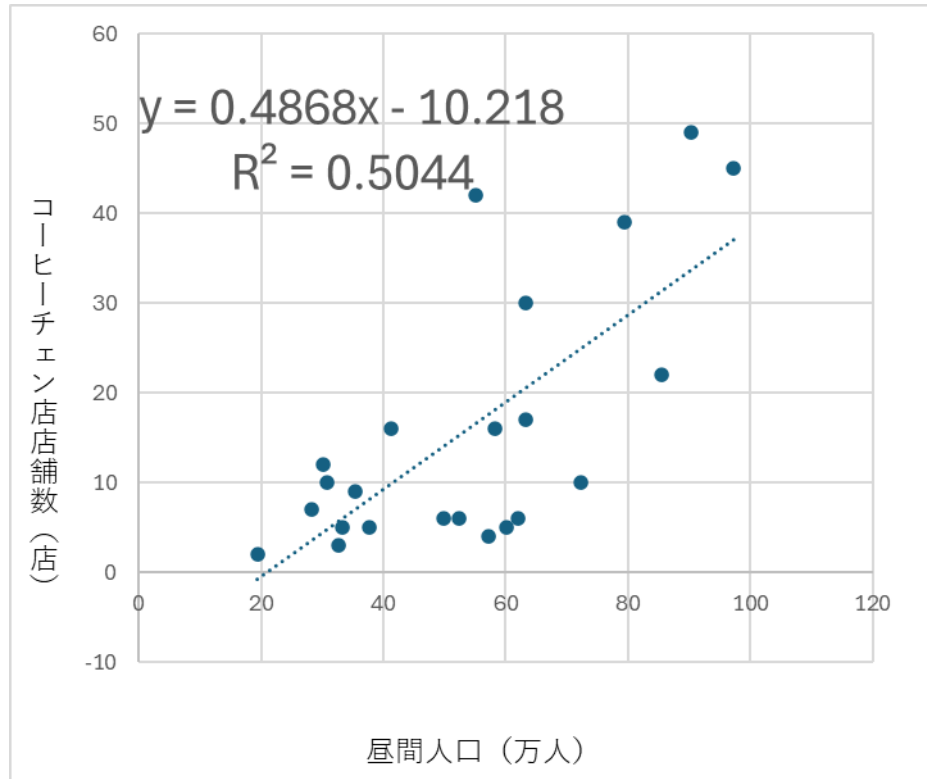
切片                  傾き                  誤差項



## ①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# エクセルでも回帰分析はできる



### 「数式の表示の仕方」

- ・データを選んで右クリック
- ・近似曲線の追加
- ・線形近似
- ・切片、グラフに数式を表示する、グラフにR-2乗値を表示する

- ・グラフで数式を表示できるのはxが一つの場合のみ。
- ・分析ツールを使えば、複数のxでも可能。

## ①データの「見え方」を知る

- 回帰分析：2つの変数の関係を捉える

# 推定した式を使うと何がわかるのか

- コーヒーチェーンの出店数が人口などのデータとどの程度関係があるのかがわかる。
- 関係がわかると、予測ができる。
- たとえば、以下の関係がわかる。

$$\text{出店数 (店)} = 0.49 \times \text{昼間人口 (万人)} - 10$$

- 昼間人口が100万人の場所には、39店あると予測できる。

## ②データ分析の基本姿勢

ー 思い込みに左右されない判断

# データ分析の落とし穴

- 「原因と結果」の関係ではないが、そのような関係に見えるパターンがある。

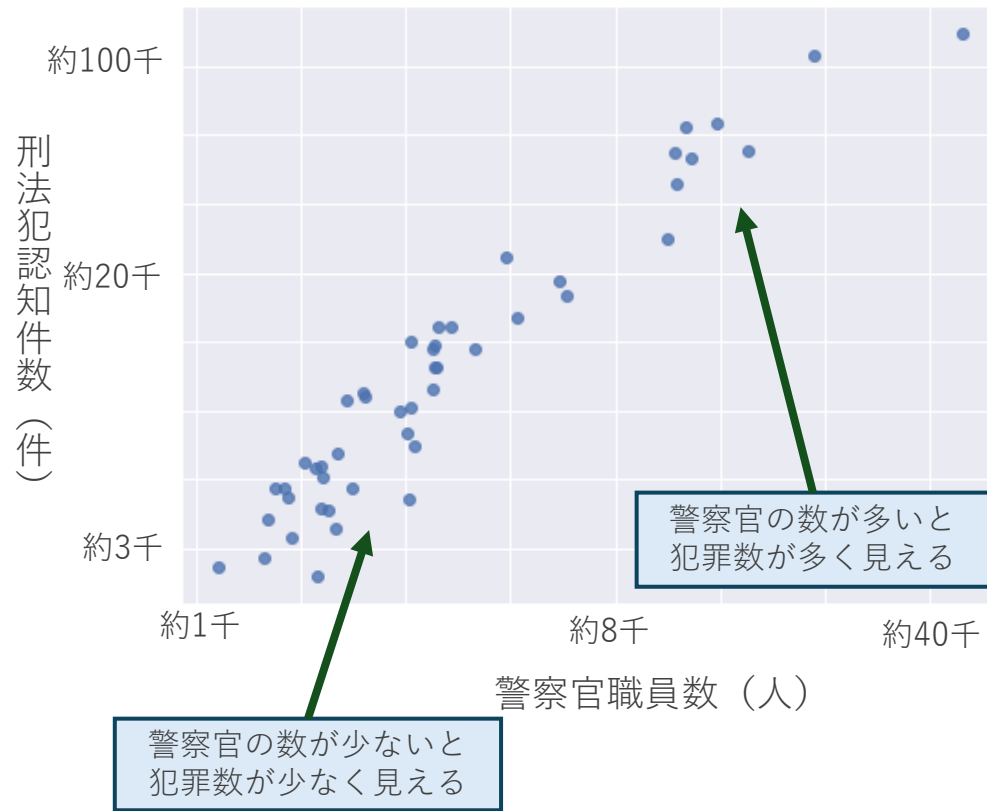
偶然	<ul style="list-style-type: none"><li>• 原因と結果の関係にあるように見える二つの要素が、たまたまそのように見えるだけのパターン</li></ul>
第三の変数	<ul style="list-style-type: none"><li>• 原因と結果の関係にあるように見える二つの要素が、それらとは別の要素によって変動しているパターン</li></ul>
逆の因果関係	<ul style="list-style-type: none"><li>• 原因と結果の関係にあるように見える二つの要素が、認識しているものと原因と結果が逆であるパターン</li></ul>

## ②データ分析の基本姿勢

ー 思い込みに左右されない判断

# データ分析の落とし穴

## 都道府県別警察官の職員数と刑法犯認知件数



刑法犯認知件数（警察庁 刑法犯に関する統計資料）

<https://www.npa.go.jp/publications/statistics/safetylife/jousei.html>

部門別職員数（総務省 地方公共団体定員管理調査（都道府県データ））

[https://www.soumu.go.jp/main\\_sosiki/jichi\\_gyousei/c-gyousei/teiin/index.html](https://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/teiin/index.html)

1. なぜ今「データ分析 × 政府統計」か？

2. データ分析は何から始めればいいの？

3. Excelを使った政府統計データの分析(演習)

原因

結果

人口

犯罪の数

警察官の数

### Memo

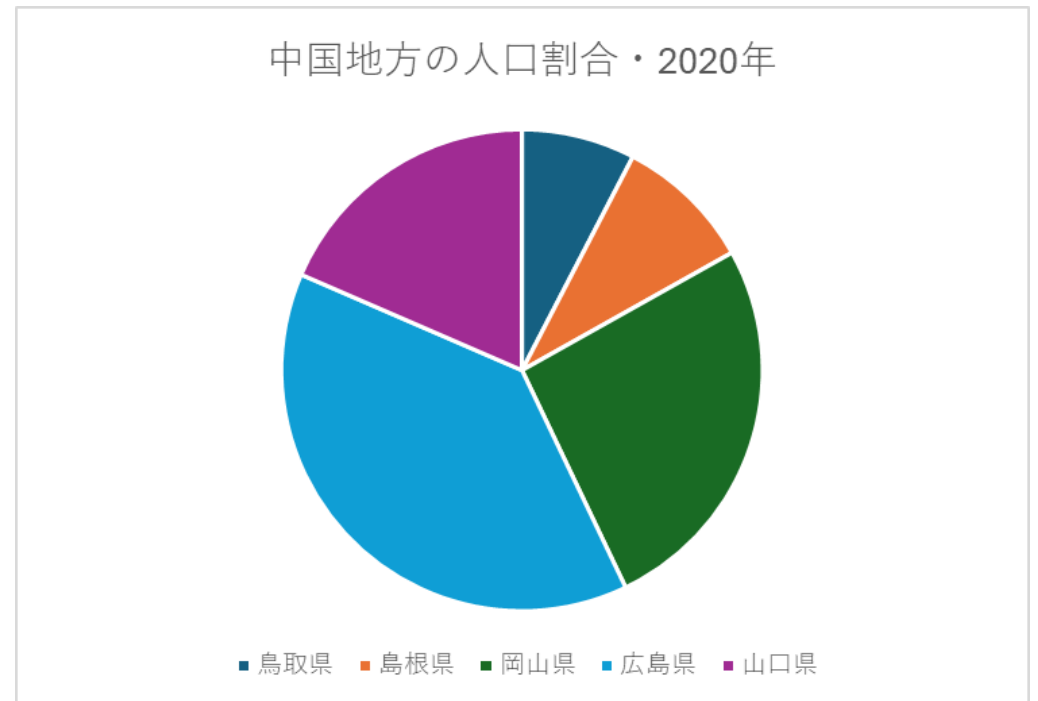
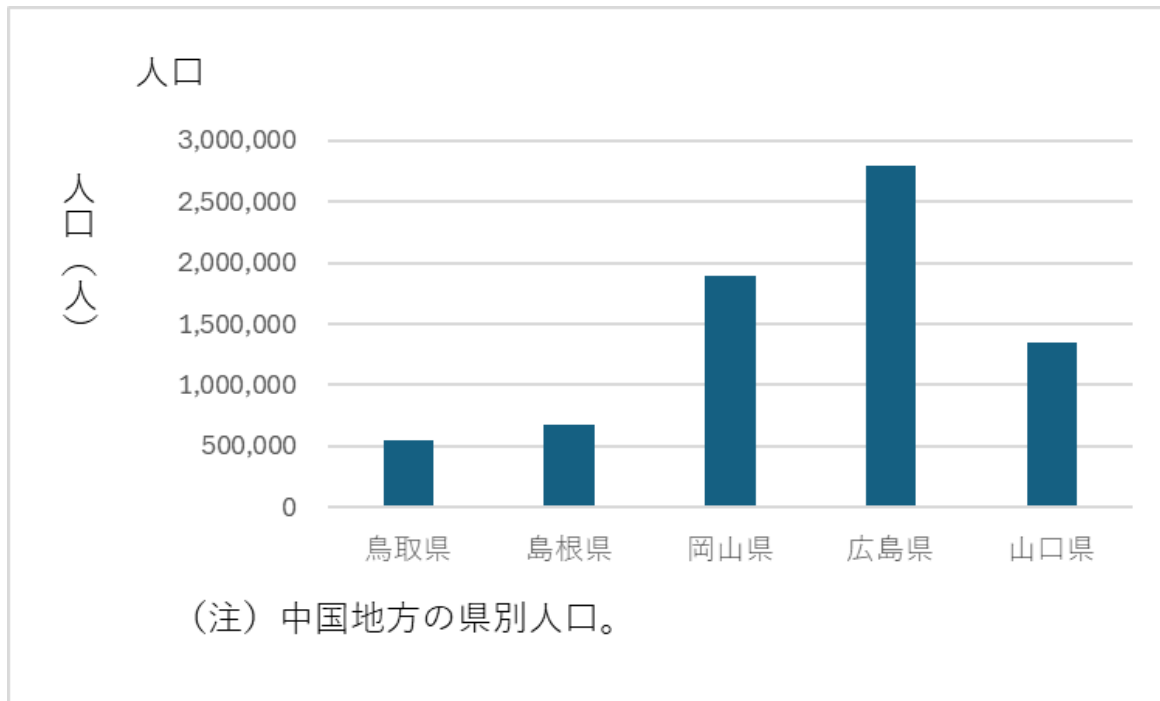
- グラフの作りは、横に行くほど警察官の数が多く、縦に行くほど刑法犯認知件数が多い作りである。
- グラフを見るだけだと、警察官の数が増えるほど、犯罪の数が増えているように見える。
- しかし、犯罪の数や警察官の数は、グラフ上に出ていない、「人口」という第三の変数に依存すると考えられる。このため、警察官の数を減らしても犯罪数は減らないと考えられる。
- なお、左のグラフは対数変換を行っているのでグラフのメモリは等間隔でない。

## ②分析の基本姿勢

ーグラフの目的に応じた使い方を学ぶ

# 棒グラフ・円グラフ

- カテゴリー別のデータには棒グラフや円グラフを使う。



## ②分析の基本姿勢

ーグラフの目的に応じた使い方を学ぶ

# 折れ線グラフ・散布図

- 時間とともに変化するデータには折れ線グラフを使う。
- 2つの変数の関係を表すには散布図を使う。

レンタカー・カーシェアリング支出額 (2010～2024年経年)

