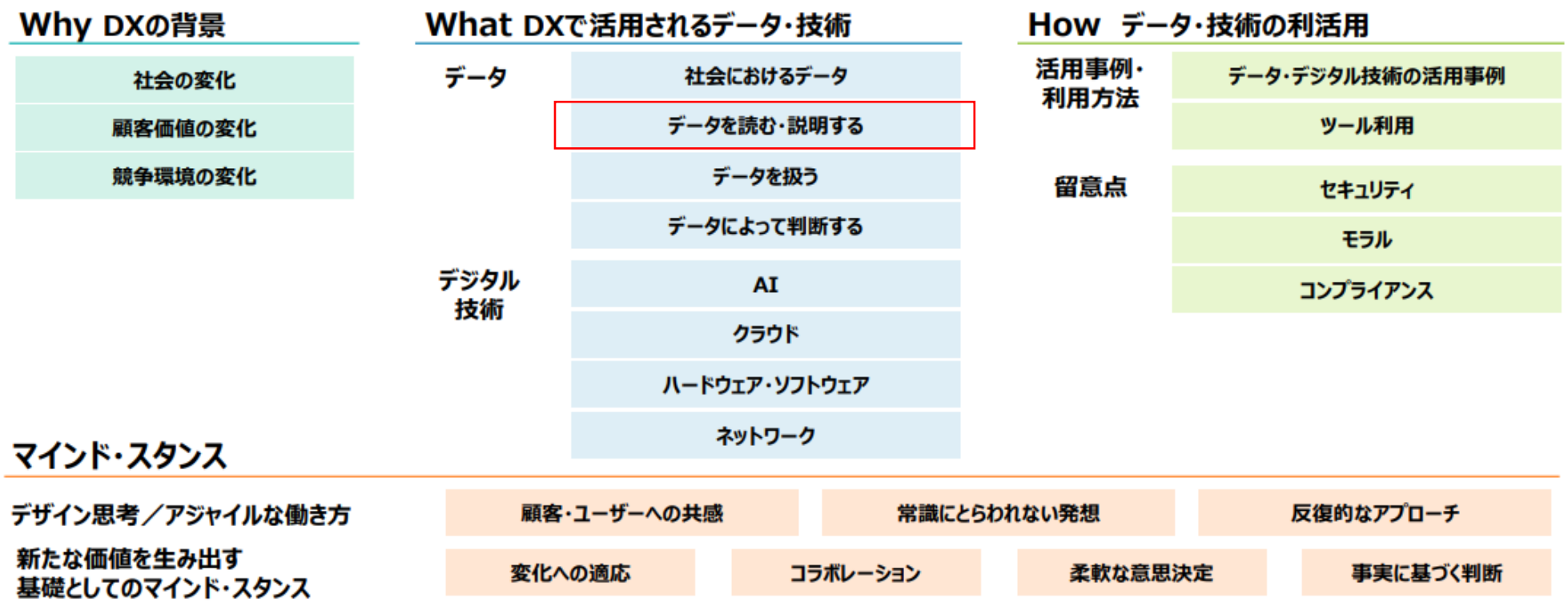
デジタルスキル標準
DXリテラシー標準
DX推進スキル標準

データを読む・説明する



作成： 2025-07-23

この教材の学習目標と学習項目

データの分析手法や結果の読み取り方を理解する。

データの分析結果の意味合いを見抜き、分析の目的や受け取り手に応じて、適切に説明する方法を理解する。

- データから得られる事実に基づいた経営・業務における意思決定を行うために、データを読み取るうえで必要な基礎的な確率・統計に関する知識や、データ同士の比較方法に関する知識を身につける必要がある。
- データから読み取った示唆を組織としての意思決定に繋げるために、結果を可視化する手法を知ることが求められる。

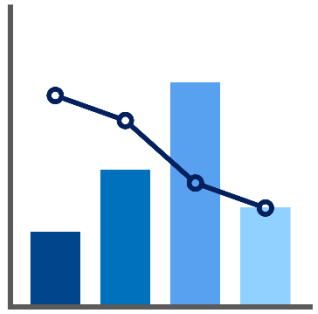
DXリテラシー標準学習項目例

- データの分析手法（基礎的な確率・統計の知識）
  - 変数、分布、関係性
  - データの種類（尺度）
- データを読む
  - 重複の発見、比較、誇張を防ぐ、ミスの特定
- データを説明する
  - データの可視化、分析結果の言語化

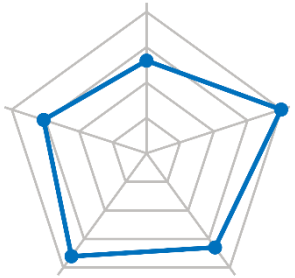
GIF追加学習項目

- テキスト分析、画像分析
- グラフの種類
- データモデリング
- 識別子、コード、統制語彙
- データ辞書
- メタデータ
- データの揺らぎ（シソーラス、オントロジー）

データのポテンシャルを引き出しましょう



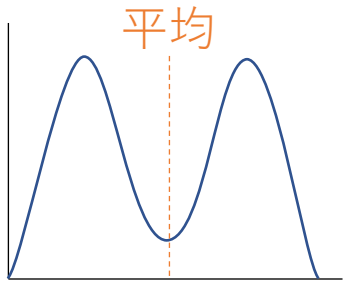
- データは組み合わせることで価値が増加します。
  - 売上高と利益率を組み合わせてみる
  - 気温と売上を組み合わせてみる
- 切り口を変えて分析すると見えてくることがあります。
  - 年代別、地域別等のクラスターに分けて人気品目を見るなど
- データ量が増えることで分析の精度が上がります。
  - 全体分析をより精緻に行うことができます
  - 全体のデータに隠れていた少数グループを知ることができます

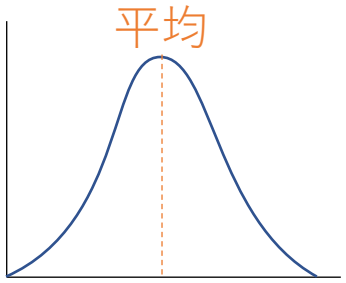


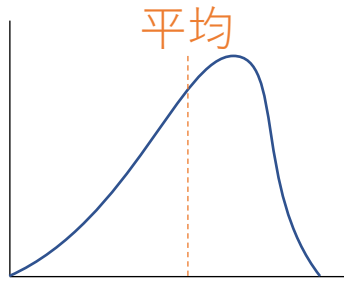
- 全体のデータと個人や個社のデータを組み合わせることで高度なサービスができます。
  - 同じ商品を買っている人のデータをもとに商品をリコメンドできます
- 生産性や効果を考えて適切なツールを導入しましょう。

読み方を知らないと誤判断が起こることがあります

- 同じ平均値でも分布が全く違うことがあります。
  - 平均の値だけでは判断できません。



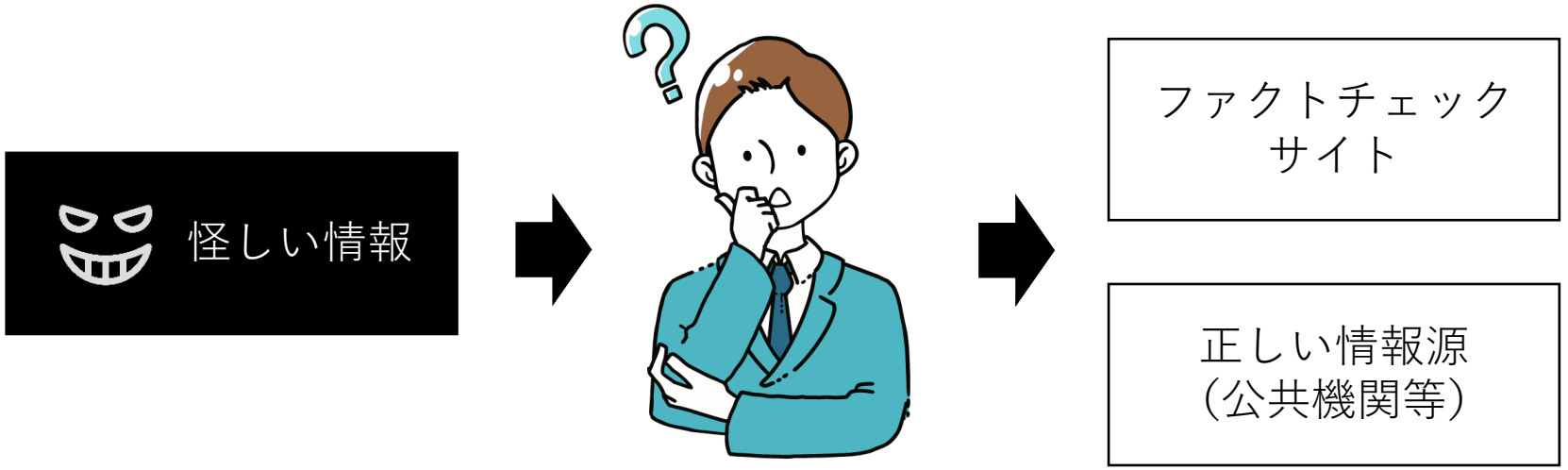




- アンケート結果は、調査対象を確認する必要があります。
  - サンプル数は十分にあるのか
  - 調査対象（年齢分布、地域分布等）に偏りはないか
- 特にAIの場合には、どのようなデータを学習したのかがAIの判断基準となるため、情報源の確認が重要です。

情報には誤情報や偽情報が入っていることがあります

- SNSの分析等を行うときには、誤情報や偽情報が入っていることを考慮する必要があります。
- 情報が怪しいと感じたときには、正しい情報を発信している情報源を確認することが重要です。
- 自組織に関する誤情報や偽情報が流れているときは、自組織のウェブサイトやSNSを使って正しい情報の提供や注記喚起を行います。



データから原石をみつけましょう

- 大量データの中の少数データを分析することでマーケットの動きを見ることができます。
  - 特定商品の売上データが上がり始めた瞬間
    - 伸び率を見ることで確認できます
  - 自由記述に意見が出始めた瞬間
    - 類義語検索などを使いニーズを明確化できます

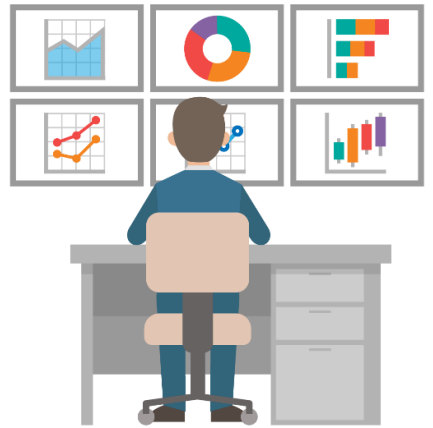
データが多くても**フィルターをかける**ことで、
大事なものを見つけることができます。

データの見せ方次第で信頼を高めることができます

- 透明化が第一歩です。
  - 企業が決算公告を自社ウェブサイトに掲載して広く伝えている会社と、掲載していない会社では信頼感が違います。
  - 悪い結果も公開することで、改善する意思が伝わります。

- 結果だけでなくプロセスを見せることが重要です。
  - 食品の製造販売プロセスや原料データ、トレーサビリティデータを見せることで信頼感が高まります。

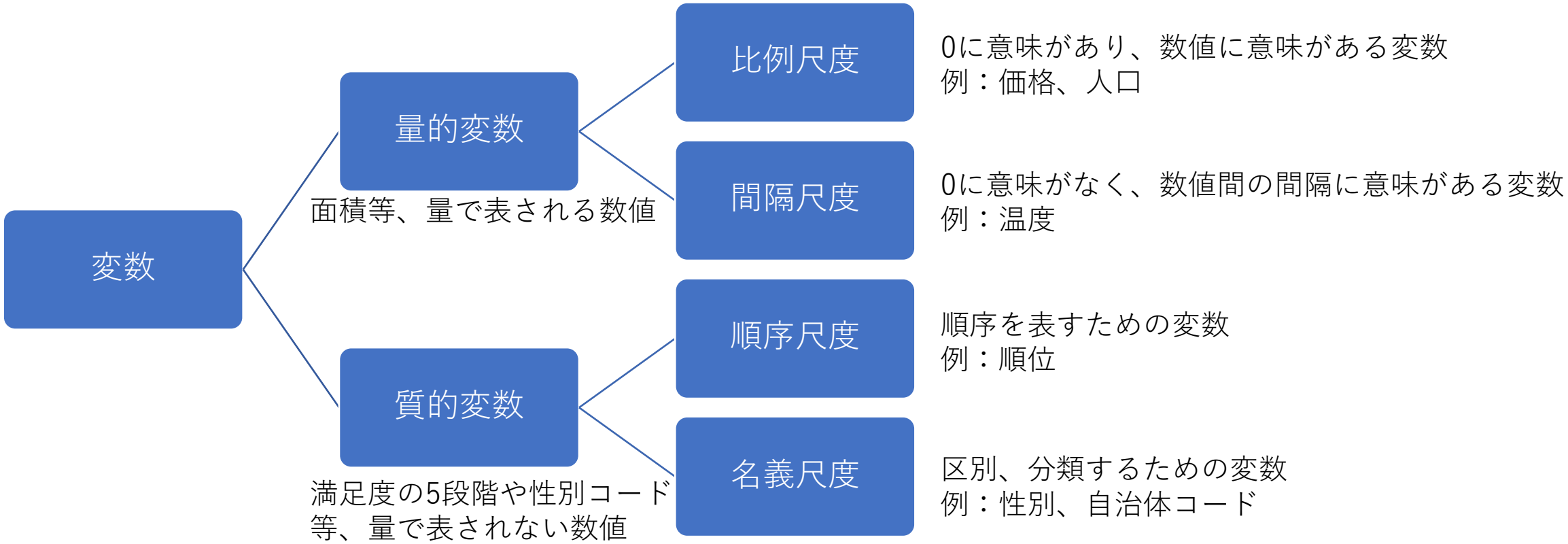
- ダッシュボードを公開する取り組みも行われています。
  - 経営等の重要な指標を示すダッシュボードを内部で活用するだけでなく、指標の一部を外部に公開する取り組みも行われています。



分析のためのデータの基礎

変数と尺度

- 統計で扱われるデータは、量的変数　質的変数と4つの尺度で分類されます。



－9点と言っても、10点満点か100点満点かで意味が違ってきます。

データの分類

- データは時間、整理方法等により、いくつかの分類があります。

データの流れによる分類

フローデータ

- 時間とともに変化していくデータであり、保存されないこともある

ストックデータ

- 蓄積したデータや記録用のデータであり、システムやサービスから参照される

整理方法による分類

時系列データ

- 時系列で取得したデータ

クロスセクションデータ

- ある時点における分類した複数の項目を集めたデータ

パネルデータ

- 同一の項目について継続的に調査して記録したデータ

システム的な分類

マスターデータ

- 複数のサービスから使われるシステムの基盤となるデータ

トランザクションデータ

- 売り上げなどの処理に関するデータであり、長期保存はされない

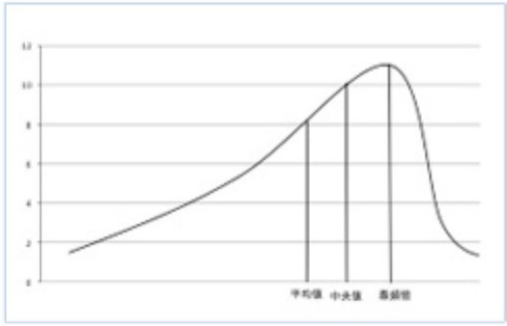
スナップショットデータ

- ある時点でのデータ

データの代表値とバラツキ

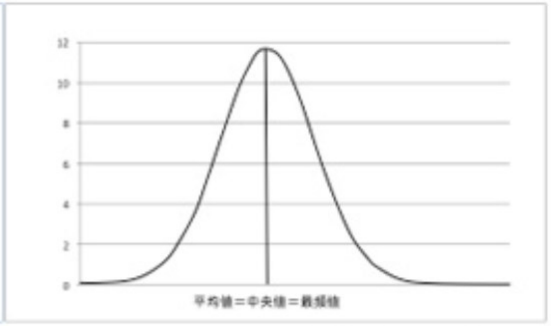
- 平均値や偏差値はデータ分析の基礎です。理解しましょう。

図　平均値、中央値、最頻値の違いＡ



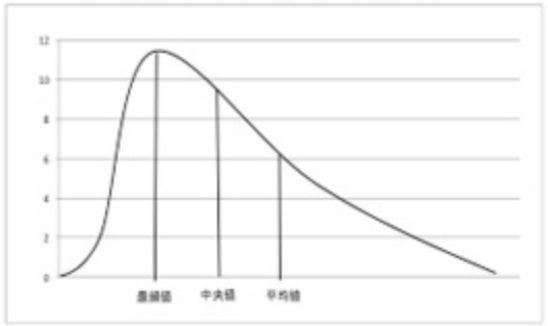
平均値＜中央値＜最頻度

図　平均値、中央値、最頻値の違いＢ



平均＝中央値＝最頻度

図　平均値、中央値、最頻値の違いＣ

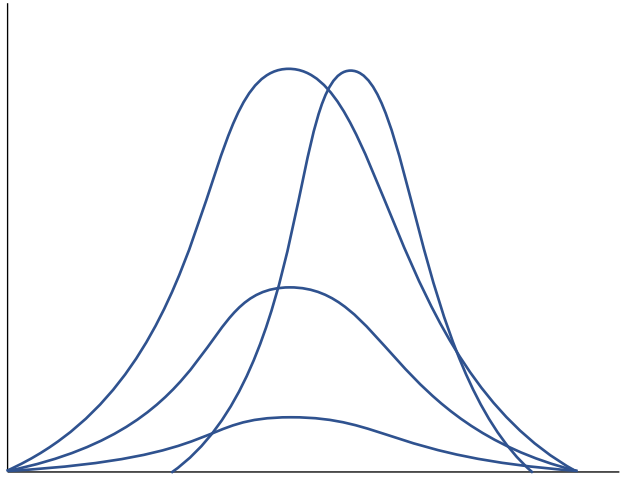


平均値＞中央値＞最頻度

出典：総務省統計局,「なるほど統計学園」

平均値
全体数値の平均の値
中央値
母集団の中央の値
最頻値
最も出現頻度の高い値

偏差
平均からの隔たりの大きさを表す値
分散
偏差を二乗して平均したバラツキ度合い
標準偏差
偏差は個々のデータごとに異なるため、データ全体として平均値からどれくらいばらついているかを示す値
偏差値
データの値を平均50、標準偏差10のデータに標準化したときの値



データのバイアス

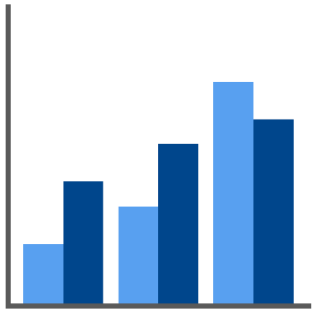
- データの取得方法によっては、データの内容に偏りがある場合があります。
  - 古いデータを使用すると、「女性は主婦が多い」といった古い考え方が反映されていることがある
  - データセンサーに特有の測定誤差が含まれている
  - データ取得者の年齢が均一でない

- このようにデータに何らかのバイアスが含まれている場合があるため、そのデータの取得方法などを確認する必要があります。

相関関係や因果関係

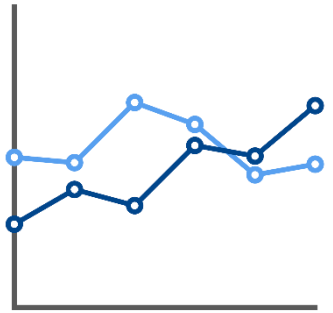
- 複数のデータの間で、値の変動が同時に生じる傾向が見られる関係を考えます。

– 相関関係
  - 複数のデータが関連していて、１つのデータが変化すれば他も変化するような関係。
    - › １つのデータが増えると他のデータも増える正の相関と、１つのデータが増えると他のデータが下がる負の相関がある



例）売り上げが伸びると利益が増大する

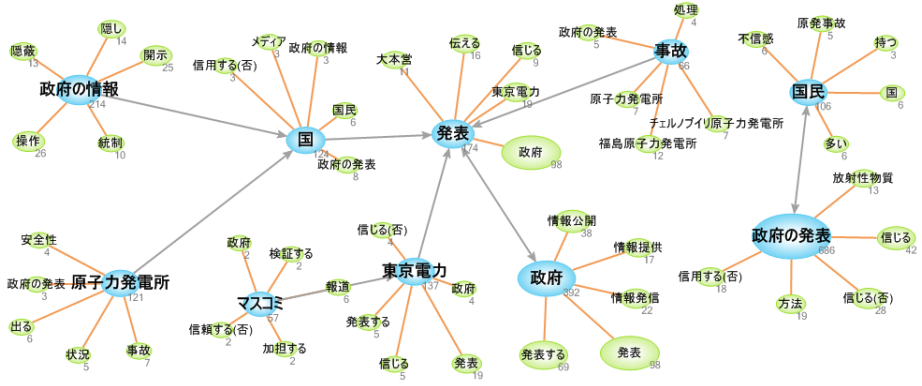
– 因果関係
  - 複数のデータの間に原因と結果の関係があること
    - › IT投資を行ったから利益が上がったのか、利益が上がったからIT投資をしているのか、数値だけでは因果関係を判断できない場合が多い



例）為替が動くと後追いで株価が変動する

テキスト分析の方法

- 文書データを解析することで、数値にはない情報を得ることができます。



  - 形態素解析
    - 文章を単語単位に切り分けます。
  - 頻出語分析
    - 単語単位に切り分けられた単語の頻出度合いを見ます
      - › よく使われる単語を見ることで、興味や注力している単語を明確化できます
      - › 文字の大きさで頻出度合いを示すワードクラウドがよく使われます
  - 共出現の相関の分析
    - 「売上」-「上がる」 「売上」-「下がる」等の相関性の分析で傾向を見ることができます
    - さらに「売上」-「上がる」 が含まれる文章の原文を確認することで深く分析できます
- 文書解析では、固有名詞や類義語、略語等の辞書の整備が重要になります。
  - 「テキスト分析」＝「テキストの分析」＝「テキストマイニング」＝「テキスト・マイニング」等

画像データ分析

- 監視カメラ、車載カメラ、衛星画像等、画像データを使う場面が増えてきています。
- 一方、画像は情報量が多いことから人の力で分析することは困難です。

- 画像処理などのツールを使って分析します。
  - 画像データの分析をした場合、その原データ、処理方法などを説明することが求められます。

データの表現

グラフの種類1

- データの分析には様々なグラフが使えます。

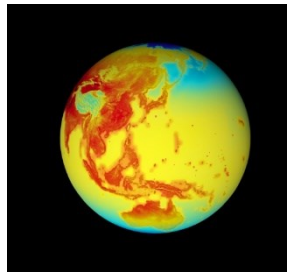
絵グラフ		: 同形の絵を並べ、量の大小を比較する。
棒グラフ		: 棒の高さで、量の大小を比較する。
折れ線グラフ		: 量が増えているか減っているか、変化の方向をみる。
円グラフ		: 全体の中での構成比をみる。
帯グラフ		: 構成比を比較する。
ヒストグラム		: データの散らばり具合をみる。
箱ひげ図		: データの散らばり具合をみる

出典：総務省統計局,「総務省統計局,「なるほど統計学園」」

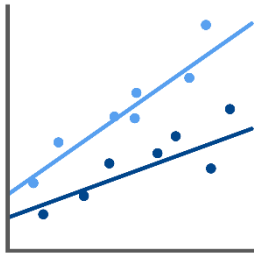
グラフの種類２

- 散布図
  - 売り上げと利益等の２つの軸で数値を分析する方法です。グラフにプロットする点の大きさや色で分類するなどで、さらに深い分析ができます。

- ヒートマップ
  - 数値を色で表すことで、視覚的に情報を表すことができます。
    - 人口密度や環境分析、画面クリック履歴の解析等でよく使われます。



- 複数のデータを組み合わせた複合グラフも用いられます。



データモデル

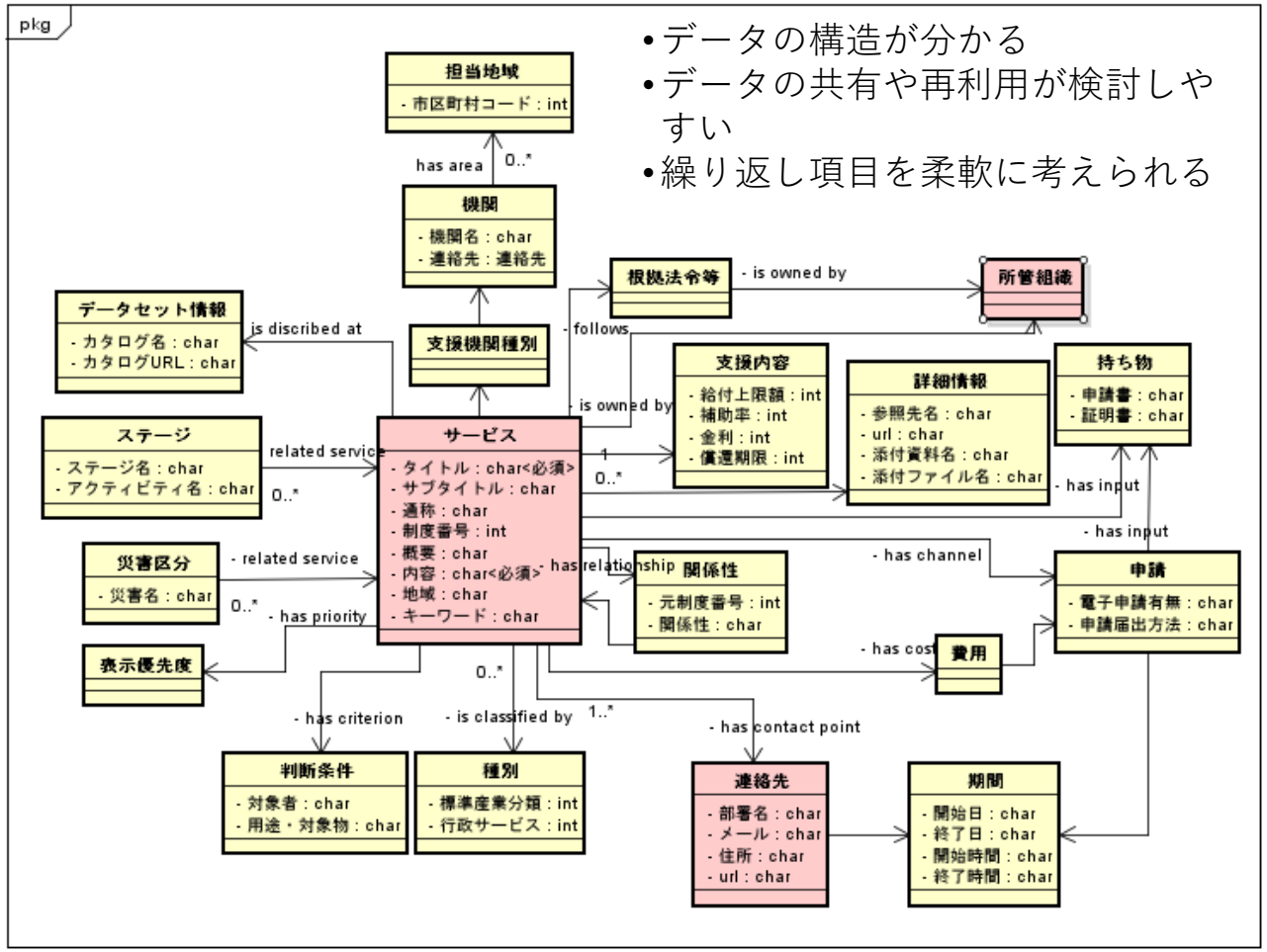
- 集計データではなく、個々のデータを読んだり、説明をするには、データがどのような要素で構成されているのか設計図を使う必要があります。

クラス図

- データを、データ対象、データ項目（属性）、操作で表し、そのデータ対象間の関係性を明示することで全体のデータ構造を示します。
  - ✓ 操作を省略したクラス図がデータモデル表記の主流になっています。

ER図

- データを、データ対象、データ項目（属性）で表し、そのデータ対象間の関係性を明示することで全体のデータ構造を示します。
  - ✓ 操作を省略したクラス図と同じ内容ですが、関係性の表記方法が複数種類あるためER図を使うことは少なくなっています。



- データの構造が分かる
- データの共有や再利用が検討しやすい
- 繰り返し項目を柔軟に考えられる

データの識別子とコード、統制語彙

- データは一意な識別子で管理することで、様々なデータをつなぎ合わせることができます。
  －法人番号を使って複数データの連携を行う等
- また、データをつなぐときに、データの分類体系が違っているとデータが効率的につなげないので、コードやコントロールド・ボキャブラリ（統制語彙）を使います。

識別子（ID）	コード	コントロールド・ボキャブラリ
• 車のナンバー等、一意に対象物を特定します。 • 駐車場、車検情報等の情報を識別子を使ってつなぐことができます。	• 産業分類等、多くの対象物を分類するために使います。 • 英数字の記号が付いていることが多いです。 • 同じコードを使っているデータは、分析や活用が容易にできます。	• 「準備中」、「開店中」のように、複数のデータで共通的に使う選択肢です。 • 同じコントロールド・ボキャブラリを使っているデータは、分析や活用が容易にできます。

データ定義の揺らぎとデータ辞書

- データの定義が異なるため、データを扱う人によって同じデータに対する理解が相違する場合があります。それを防ぐために、組織横断でデータ辞書を作る必要があります。
  - 「世帯」は、寮や老人ホームは含む定義と含まない定義があります。どちらの定義を使うかで集計値に差が出てきます。
    - 世帯の法的定義は、「住居及び生計を共にする者の集まり又は独立して住居を維持し、若しくは独立して生計を営む単身者」で寮等を含みます。一般の調査では、寮等を含まない場合が多いです。
  - 部門や企業によって、在庫消込の定義が「出荷時」「納品時」と基準が違う場合もあります。
- また、データ項目名や内容表記に、「企業名」「法人名」「会社」等の揺らぎがある場合もあります。類義語を管理するシソーラスや、対象に対する概念を管理するオントロジー等の活用検討を行います。

データ管理や検索のためのメタデータ

- メタデータは、「データのためのデータ」といわれる検索や管理用のデータです。デファクト標準のDCATが世界で普及しています。

タイトル	説明	作成者	作成日	備考
＊＊入門	＊＊＊＊＊＊	田中一郎	2020-12-01	
＋＋の手引き	＊＊＊＊＊＊	佐藤次郎	1987-02-03	廃刊
＆＆の本質	＊＊＊＊＊＊	近藤太郎	2020-01-05	

- 管理用の情報を正しくつけることはデータ作成の現場では手間が増えるため、嫌がられることがありますが、データの管理や再利用が簡単になり、中長期的にメリットがあります。
  - 現場の理解を深めていくことが重要です。

データの読み方と説明の仕方

データを読むとは

- 誰かがデータを持ってきたときに読みこなす力が必要です。

- データがきちんとした基礎知識に基づいて整理されているかどうか。
  - 関連したデータはないのか（探しきれているか）
  - 原データの品質は確保されているか
  - データの処理方法は問題ないか
- 都合の良いデータだけを持って説明にくることが多いので、説明のスキマをきちんと確認しましょう。
  - 「８割の人が賛成しています」→「２割の人はなぜ反対しているのだろうか？その理由はわかりますか？」

データを読む

- 傾向をつかむ
  - その瞬間のデータをスナップショットのデータと言います。データ変化の方向性、速度、加速度等の傾向を考えることが必要です。
    - マーケティングでは、０が１になる瞬間が重要と言われます
- データや事象の重複に気づく
  - 重複したニーズを知る
    - 多くの人が同じ指摘をしていると言うことは、重要な情報です
  - 重複したデータを省く
    - 同じ人による重複回答を除き、データ分析の正確性を高めことが必要です
    - SNS分析では再送や重複投稿されたデータがありますが、フィルター機能などを使用して適切に処理することが必要です
      - 再送されることには数的な意味合いがあり、それも一つの結果として評価します
      - 一方で、再送が多いとテキスト分析において異なる結果が出るため、再送を除いた分析を行うことで埋もれた意見を発見しやすくなります

データを読む

- 条件をそろえて比較します。
  - 対象地域やコードをそろえた上でデータの比較を行うことで正しい分析や判断ができます。
    - 前提条件が補正できるときには補正する
    - 前提条件がそろわない場合には、その違いを明記する
- データの説明における誇張表現を見抜きます。
  - データが強調されるようにグラフの縦横比を変えたり、軸の途中を省略したりすることで、視覚的錯誤が起こります。
    - 0-100%の軸から80-100%の軸にすると差を強調できる
  - 軸の目盛の確認などで誇張説明されたグラフを正しく読む必要があります。
- 集計ミス・記載ミスを特定したらフィードバックします。
  - 複数データの不整合などを見つけたときには、データ作成者にフィードバックします。（DataOps）
    - 次回以降のデータ品質向上につなげていく。

データを説明するとは

- データを正しく理解してもらう必要があります。
- データを説明するには、内容のサマリーと根拠となるロジックや原データが必要です。
  - 原データリスト
    - 収集対象
    - 収集方法
    - 最終更新時期
  - 処理プロセス
    - マッシュアップの方法
    - クレンジングの方法
    - コードなどのマッピング
    - 集計の方法（類推や比例配分など）
- 適切なグラフや特徴的な数字を使って説明します。
- 必要な場合には、グラフだけでなくポイントを文章で表現します。
  - プレゼンテーションではアニメーション等を使うことで強調できます

まとめ

- データは意思決定の基盤であり、正確に理解する必要がある
- データは間違っている可能性もあるので、その可能性を知っておく必要がある
- データの読み方を知っていることで、新たな発見ができる
- データを説明するときには、その根拠を明示する必要がある
- 悪いデータでもきちんと説明することで信頼を得られる

改訂情報

- 2023-03-31
  - デジタル庁がGIFアカデミーとして公開
- 2025-07-23
  - GIFアカデミーの資料を元に以下の内容を追記し、他教材とともにシリーズ化
    - バイアスデータの追加