

# データ利活用 ガイドライン

Version 1.0.0

2024年3月

東京都デジタルサービス局



# 目次

## はじめに

- 1 ガイドラインの背景および目的 P.4
- 2 ガイドラインの構成 P.7

## 企画

- 1 問題を特定し仮説を構築する P.13
- 2 仮説検証のための計画を立てる P.17
- 3 ワークシートを活用して企画する P.20

## 収集

- 1 データを収集する (Data) P.29
- 2 データ収集における留意事項 P.32

## 利活用

- 1 データを分析する P.41
- 2 分析結果を考察する P.53
- 3 利活用の実践例 P.56

## チェックリスト

- 1 利用許諾チェックリスト P.60
- 2 データ品質チェックリスト P.61

# はじめに

はじめに編では、データ利活用が求められる社会的な背景や本ガイドラインの目的・構成・利用シーンなど、ガイドライン本編を読む前に知るべきことを紹介します。

- ① ガイドラインの背景および目的 

---

 **P.4**
- ② ガイドラインの構成 

---

 **P.7**

## 1 ガイドラインの背景および目的

### 1. 本ガイドライン策定の背景 データ利活用のメリット

先行きが不透明で変化の速度が速い現在の社会では、人々が抱える課題や価値観も多様化・複雑化し、ニーズを把握することがより難しくなっています。このような環境の中で適切な施策を打ち、品質の高い行政サービスを提供するために、データの利活用が期待されています。日々の業務の中でデータに基づいて考え判断することで、ニーズに合った施策の立案や、施策の定量的・客観的な検証を行うことが可能になります。

他にも、データを可視化や分析することで複雑な情報を整理し課題を発見できること、業務で活用しているデータの公開によって都民のQOL向上につながる新しいサービスが創出されることなど、様々な効果があると言えるでしょう。

#### データ利活用のメリット



## 1 ガイドラインの背景および目的

### 2. 本ガイドラインの目的

データ利活用が重要とは言うものの、具体的な作業手順や留意事項がわからない方も多いのではないのでしょうか。

本ガイドラインの目的は、職員の皆さんが自分の業務でもデータを活用する効果がありそうと感じてもらい、実際に取り組んでもらうことにあります。

ひとりひとりが理解を深め実践することで、組織がデータを利活用しやすい環境となり、行政サービスの品質を向上させることにつながります。

本ガイドラインでは、データ利活用の進め方に沿って各フェーズでの主な作業内容や注意事項を説明しています。データを利活用するとは何なのか、考えながら読んでいきましょう。



## 1 ガイドラインの背景および目的

### 3. 本ガイドラインの位置づけ

2022年3月、品質の高いデジタルサービスを安定的かつ永続的に提供し、都政のQOS向上に寄与するために

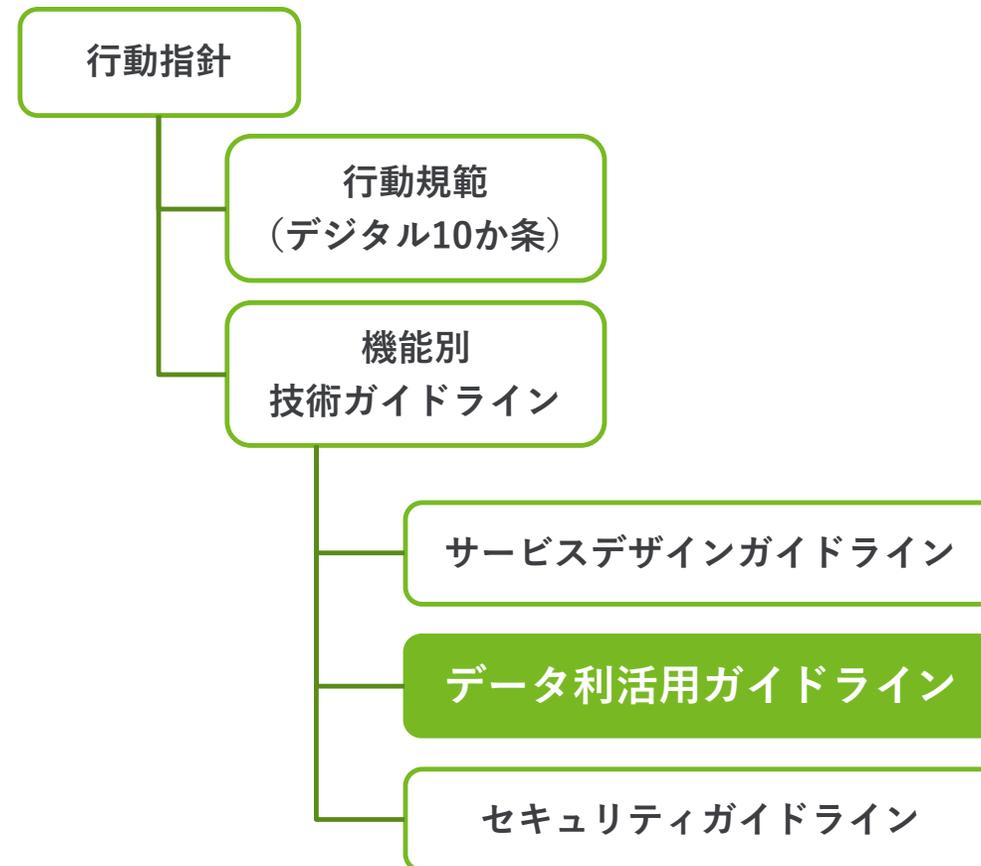
「東京都デジタルサービス開発・運用に係る行動指針」が策定されました。これは、デジタルサービスの開発・運用に携わる全ての職員等が遵守すべき、基本的な価値観などを示しています。

この行動指針は、「行動規範（デジタル10か条）」と「機能別技術ガイドライン」から構成されます。

本ガイドラインは「データ利活用」をテーマとした機能別技術ガイドラインであり、行動規範（デジタル10か条）の実践にあたって必要な考え方や行動を説明するものです。

行動指針・行動規範と合わせて読むようにしましょう。

#### 行動指針の構成



## 2 ガイドラインの構成

### 1. ガイドラインを構成するデータ利活用のフレーム PPDAC

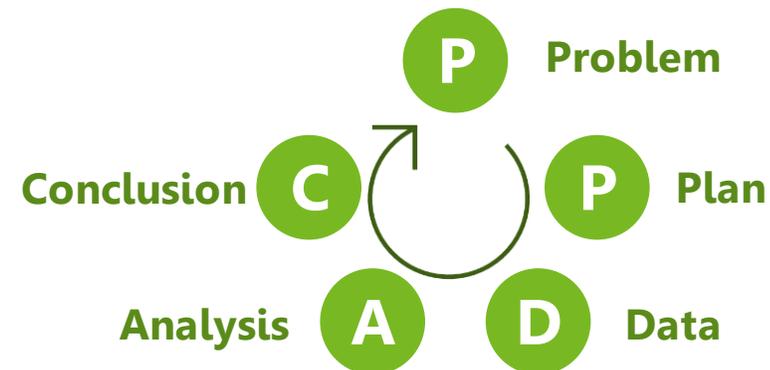
イメージがわきやすいよう、本ガイドラインはデータ利活用の進め方にあわせて構成しています。

データ利活用の進め方には様々なパターンが存在しますが、本ガイドラインでは「PPDAC」※1というフレームワークの活用を推奨します。PPDACは問題・計画・データ・分析・結論の5ステップを持ち、特に問題や原因仮説を検証するデータ分析において有用です。

データ利活用は目的ではなく手段であり、PPDACは「なぜデータを利活用するのか？」という目的を見失わないように整理されています。フレームワークに沿うことで手順が明確化し理解しやすくなるほか、一定の品質も担保できるため、本ガイドラインではPPDACを活用したデータ利活用を前提として説明します。

※1：PPDACはデータを利用した問題解決のフレームワークとして一般的にも使われており、総務省統計局のData StaRt(巻末付録：リンク一覧①)などで紹介されています

#### PPDACのフレームワーク



#### ワンポイント

- 本ガイドラインではPPDACを活用した「分析」を前提としますが、データ利活用には「分析」以外にも「可視化」や「予測」など様々なパターンがあります
  - 可視化**：データや情報をわかりやすいようにビジュアルライズし共有する  
→Analysis編で事例を紹介
  - 予測**：データを分析したパターンを元に、将来の値・事象を精度高く計算する  
→高難易度なため本ガイドラインでは割愛

## 2 ガイドラインの構成

### 2. PPDACに基づくガイドライン全体の構成

本ガイドラインは5つの区分で構成されており、企画編・収集編・利活用編はPPDACの5つのフェーズに紐づく内容となっています。また最後には、利用許諾取得やデータ品質管理で活用できるチェックリストを紹介しています。

#### PPDACの5つのフェーズ

解決すべき問題と原因仮説、つまり分析の目的と問いを設定する

仮説を検証する分析方法や必要なデータの収集方法を計画する

分析に必要なデータを収集する

収集したデータを用いて、計画にもとづいて分析作業を実行する

分析結果から仮説の妥当性を判断し、検証された内容を業務に反映する

**P**roblem

**P**lan

**D**ata

**A**nalysis

**C**onclusion

#### 本ガイドラインの構成とトピック

はじめに

企画編

収集編

利活用編

チェック  
リスト

- 本ガイドラインの背景・目的や構成
- Problem/Planフェーズでの実施事項と注意事項
- データ利活用ワークシートの概要と活用方法
- Dataフェーズでの実施事項と注意事項
- 収集したデータの取り扱い
- Analysis/Conclusionフェーズでの実施事項と注意事項
- 主な分析の観点と特徴
- 利活用許諾とデータ品質の確認事項

## 2 ガイドラインの構成

# 2. PPDACに基づくガイドライン全体の構成

### 目次別のメッセージ一覧

区分	章	見出し	記載内容
はじめに	1	ガイドラインの背景および目的	<ul style="list-style-type: none"> <li>本ガイドラインの位置づけと目的、利用シーン</li> </ul>
	2	ガイドラインの構成	<ul style="list-style-type: none"> <li>PPDACの一連の流れと各フェーズでの実施事項</li> <li>PPDACを踏まえた本ガイドラインの構成</li> </ul>
企画編	1	問題を特定し仮説を構築する	<ul style="list-style-type: none"> <li>分析着手前に問題を設定することの必要性</li> <li>問題特定と仮説立てのステップと考え方</li> </ul>
	2	仮説検証のための計画を立てる	<ul style="list-style-type: none"> <li>分析着手前に、PPDACのData/Analysis/Conclusionを具体的に決めることの必要性</li> <li>計画策定における検討事項</li> </ul>
	3	ワークシートを活用して企画する	<ul style="list-style-type: none"> <li>PPDACワークシートの概要</li> <li>企画工程におけるワークシートの活用場面と作成手順</li> </ul>
収集編	1	データを収集する	<ul style="list-style-type: none"> <li>データ収集時の注意事項とその必要性</li> </ul>
	2	データ収集における留意事項	<ul style="list-style-type: none"> <li>データの収集において注意すべき3つのトピック（利用許諾取得、品質管理、標準化）</li> </ul>
利活用編	1	データを分析する	<ul style="list-style-type: none"> <li>データ利活用において有効な分析パターンと使い分けのイメージ</li> </ul>
	2	分析結果を考察する	<ul style="list-style-type: none"> <li>結論の導出や施策の提案を行う必要性</li> <li>分析結果から結論を導出するステップ、導出後のアクションや注意事項</li> </ul>
	3	利活用の実践例	<ul style="list-style-type: none"> <li>データ分析から施策の提案までの具体的な実践例</li> </ul>
チェックリスト編	1	利用許諾チェックリスト	<ul style="list-style-type: none"> <li>利用許諾チェックリストの利用シーンと目的、主なチェック内容</li> </ul>
	2	データ品質チェックリスト	<ul style="list-style-type: none"> <li>データ品質チェックリストの利用シーンと目的、主なチェック内容</li> </ul>

## 2 ガイドラインの構成

### 3. 本ガイドラインの利用シーン

本ガイドラインは職員が業務やプロジェクトのなかでデータ利活用を企画し実行する場面を想定しています。例えば業務におけるデータ利活用方法の検討や、実行予定の施策と課題の整合性検証などの場面が想定されま

す。システム開発事業に参与しているかどうかは関係なく全職員が対象であり、本ガイドラインでは自分の業務においてデータを取り扱う際に押さえるべき要点を説明しています。

また、末尾の品質チェックリストにはシステムの要件定義や設計など、将来的に組織内で利活用されうるデータを生成する場面も想定されています。利活用されやすいデータのために何に気を付けるべきか、理解と確認に活用できます。

#### 利用シーン

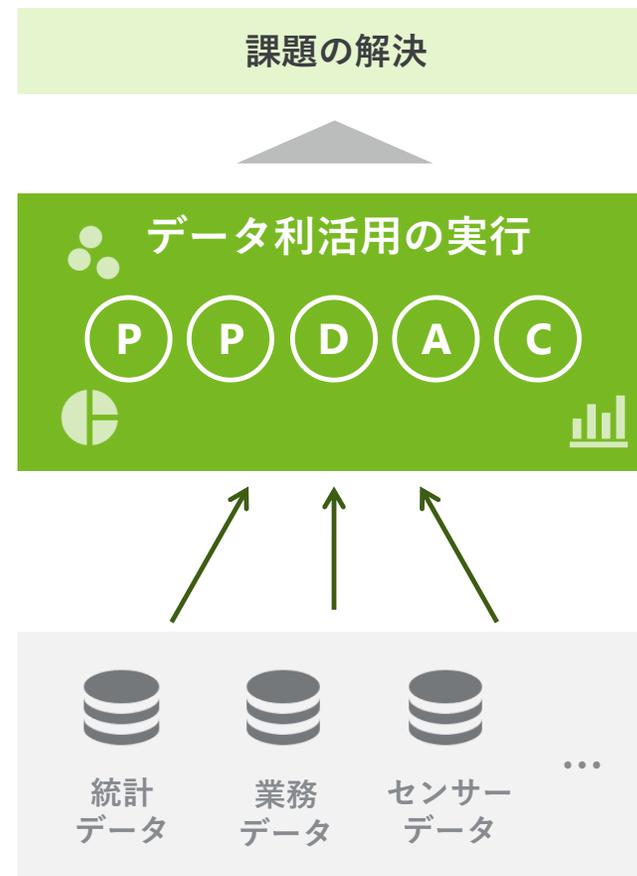
##### 1 企画・実行するとき

業務品質を上げるにはどんなデータと利活用が必要かな？



##### 2 データを設計するとき

使いやすいデータを作るために何を意識すればいいんだ？



## 2 ガイドラインの構成

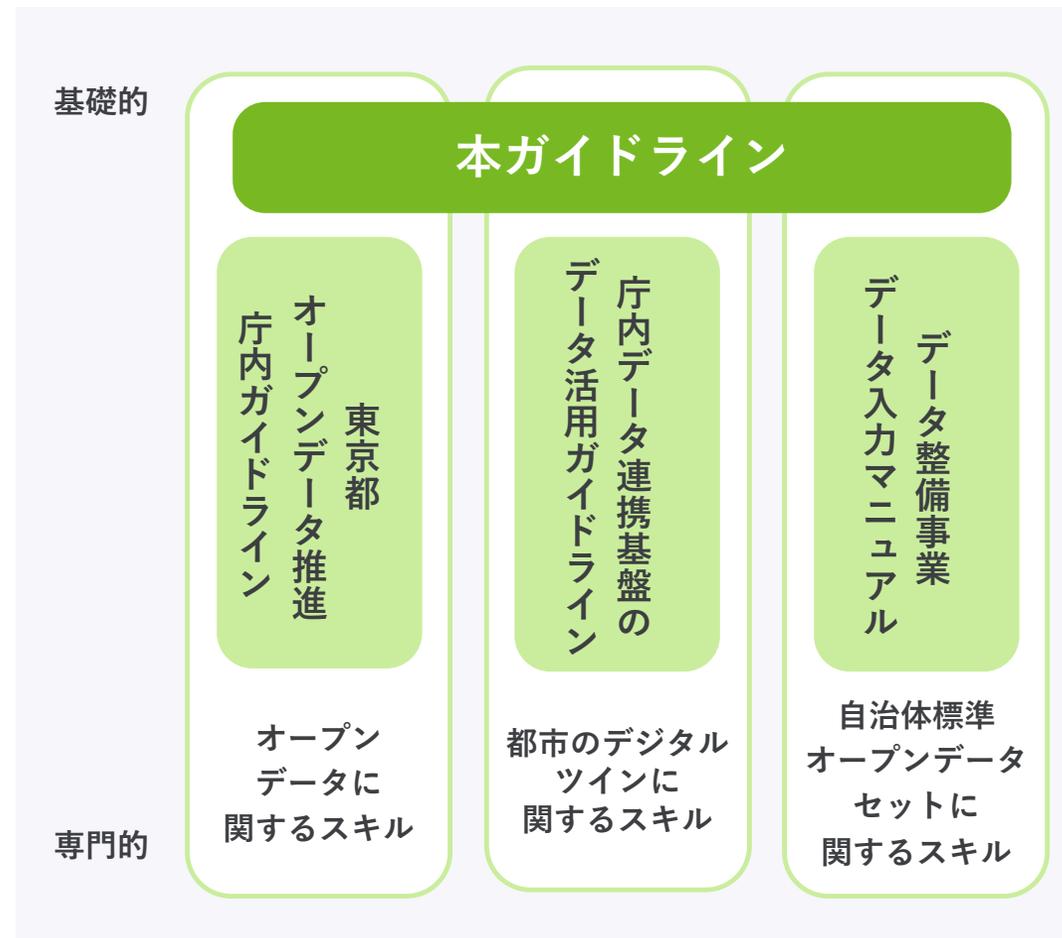
### 4. 関連する手順書・マニュアルなど

都には、本ガイドライン以外にもデータに関する手順書・マニュアルなどが存在します。これらは、付録「ガイドラインに係る手順書\_マニュアル等」で整理しているとおりです。

本ガイドラインでは、データを利活用するための基礎的な事項を整理しています。

デジタルサービス局のデータ利活用事業では、行政データのオープンデータ化（オープンデータカタログサイト）、地理空間データの庁内活用（デジタルツイン）、官民のデータ連携（東京データプラットフォーム）の事業を行っております。各事業の利用については、各事業から提供している手順書やマニュアルなどを参照してください。

### 本ガイドラインの位置づけ



# 企画

企画編では、PPDACサイクルのProblemフェーズとPlanフェーズについて順に説明します。

データ利活用を行う目的を定め、その目的から必要なデータや分析手法を計画立てることはPPDAC全体のフェーズをスムーズに進めるために重要です。

Problem/Planフェーズで決めた内容が後続フェーズに与える影響を考えながら読み進めましょう。

また本ガイドラインではPPDACの実践を支援するツールとして「PPDACワークシート」も用意しています。その概要と使用方法も紹介しますので、具体的な作業を自分の周りの業務でイメージしてみてください。

- |   |                 |       |      |
|---|-----------------|-------|------|
| 1 | 問題を特定し仮説を構築する   | ————— | P.13 |
| 2 | 仮説検証のための計画を立てる  | ————— | P.17 |
| 3 | ワークシートを活用して企画する | ————— | P.20 |

1 問題を特定し仮説を構築する

# 1. Problemフェーズとは

データ利活用は、問題解決のための手段です。そのため、問題が明確でなければその結果を有効に活用することができません。その解くべき問題を設定するフェーズがProblemフェーズになります。このフェーズで明確な問題を設定することで、後続のフェーズでも適切に仮説や計画を立てることができます。

Problemフェーズは、「問題の明確化」と「原因仮説の構築」の2つのステップで構成されます。

「問題の明確化」では、あるべき姿と現状のギャップから解決すべき事象を具体化します。それを受け「原因仮説の構築」ではこの事象がなぜ発生しているのかを検討し、原因について仮説を立てます。

## Problemフェーズのステップ



1 問題を特定し仮説を構築する

2. 問題の明確化

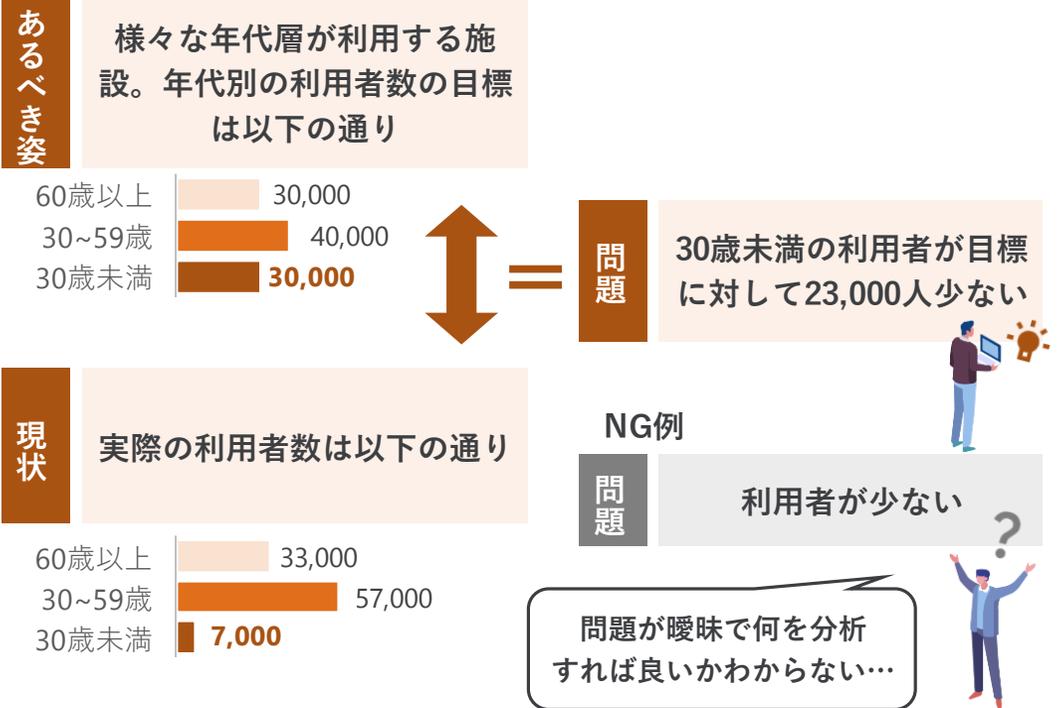
「問題の明確化」では、あるべき姿と現状のギャップから解決すべき問題を具体化します。

ここでは、あるべき姿・現状・問題の3つを具体的かつ定量的に把握することが重要です。そうすることでこの後の原因仮説や分析計画も具体的に作成できるようになります。例えば、右図のように「何が、どの程度、どんな状態なのか」まで具体的に整理することが必要です。

「利用者が少ない」のような曖昧な問題を設定してしまうと、分析計画を立てることが難しくなってしまいます。

なお、あるべき姿・現状・問題の3つはどの順番から検討し始めても、最終的に整理できれば問題ありません。あるべき姿については、組織として定めている目標やビジョンも参考にしてみましょう。

「問題の明確化」の具体例



**ワンポイント**

定量的かつ具体的な問題の設定は、後続の作業を行いやすくするだけでなく、関係者との間での認識のズレを予防することにもつながります。そのため、上記のような情報は文書化しチーム内で共有することが重要です。

はじめに

企画

収集

活用

チェックリスト

1 問題を特定し仮説を構築する

3. 原因仮説の構築

「原因仮説の構築」では、問題の主たる原因に関する仮説※1を立てます。ここでは次の2点が重要です。

1つ目は、原因を幅広く論理的に考えることです。具体例としては、右図のようにロジックツリー※2により問題の構成要素や対立構造などで分解して明確化します。

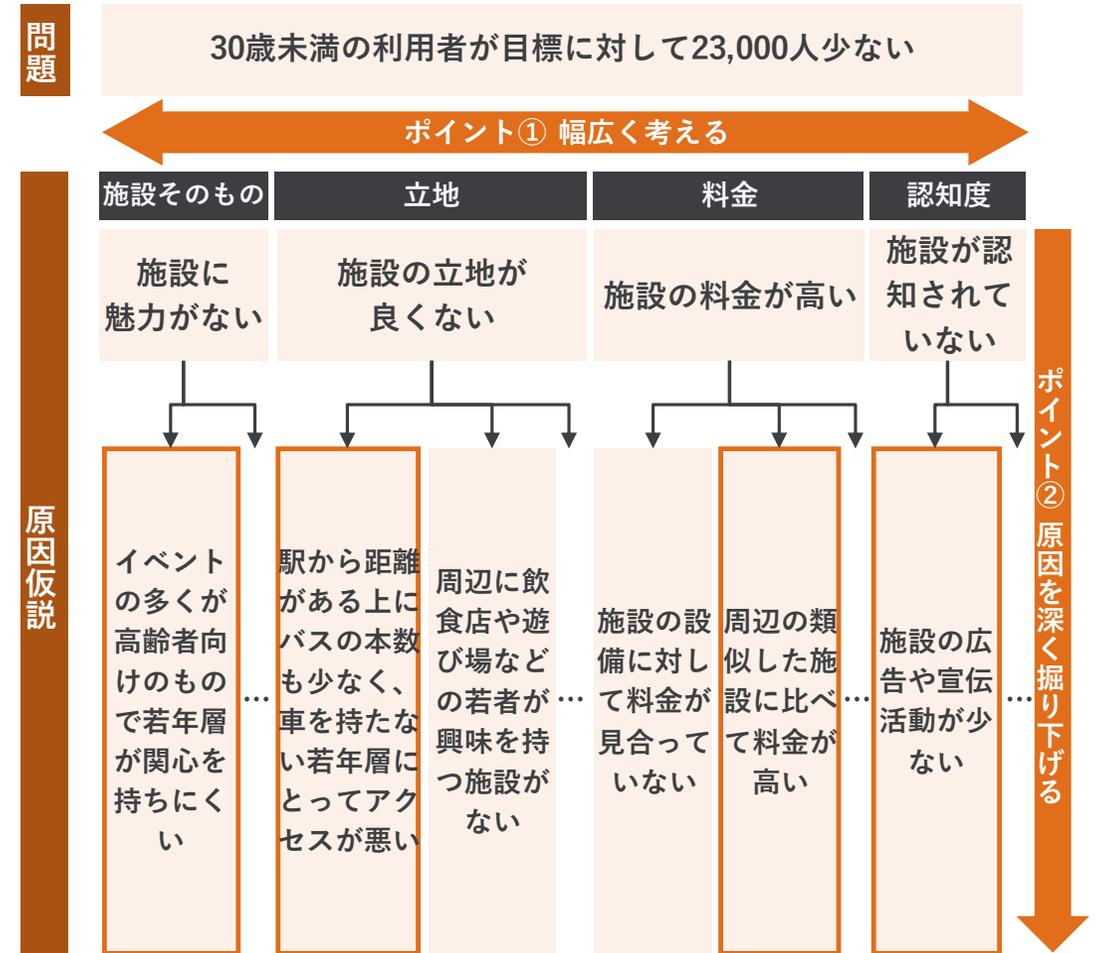
2つ目は、「なぜ」を繰り返して原因を深く掘り下げることです。そうすることで原因がより本質的になり、それを基に検討する解決策も効果的になります。

上記2点に留意して原因を洗い出せたら、そのなかから問題への影響が大きいと思われるものを選定します。ここから先のフェーズのなかで選んだ原因仮説の真偽を検証していくためこの時点では正確な根拠は不要です。

※1：仮説とは、その時点の情報から導くことができる仮の答えを意味します。仮説を立てることで検証すべきことが絞られ、効率的に分析を進めることができます。この仮説を分析を通して検証していくことになります。

※2：ロジックツリーとは、問題を構成する要素を木の枝のよう分解して書き出すことで、原因や解決策を論理的に導くためのフレームワークの一種です。

「原因仮説の構築」の具体例



洗い出した仮説のうち問題への影響が大きいものを選定する

はじめに  
企画  
収集  
利活用  
チェックリスト

1 問題を特定し仮説を構築する

**COLUMN** **PPDAC以外の分析アプローチ**

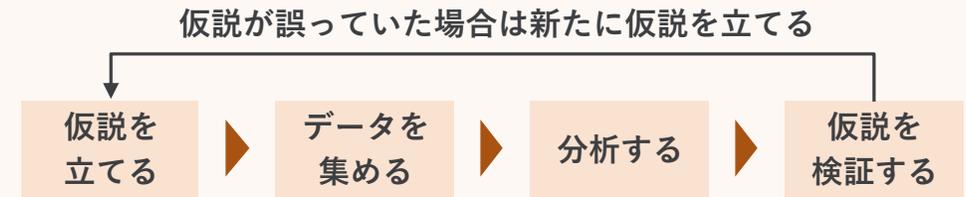
分析アプローチには検証型と発見型の2つがあります。

検証型はPPDACサイクルのようにまず原因仮説を立て、それを検証するために分析を行うというアプローチです。本ガイドラインではこちらのアプローチを適用しています。分析することが決まっているため比較的短時間で実施できるうえ、仮説が誤っていても検証結果が新しい仮説を立てるうえでのインプットになるため無駄も少ないです。

発見型はデータを様々な観点から分析して傾向やパターンを見つけるアプローチで、データマイニングなどが該当します。意外な気づきを得られる可能性がある反面、データを多角的に分析するために時間やスキルを要したり、新たな発見が得られない等のリスクがあります。

検証型

本ガイドライン対象



発見型



発見型アプローチ（データマイニング）の例

- スーパーで顧客の購入データを分析したところ、ビールとおむつが一緒に購入されるという傾向が分かった



## 2 仮説検証のための計画を立てる

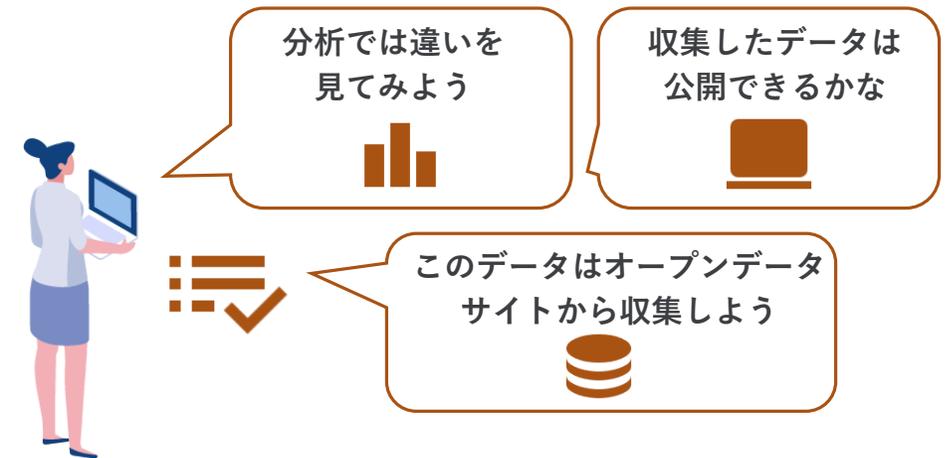
### 1. Planフェーズとは

Planフェーズでは、Problemフェーズで設定した原因仮説を検証するための計画を立てます。

目的に対して適切な分析を行うためにも事前に計画を立てることはとても重要です。仮に無計画にデータ収集やデータ分析に着手してしまうと、目的を見失ってしまい不適切な手法で分析してしまったり、分析手法に対して必要なデータが収集できない、といった非効率や無駄が生じてしまいます。

計画を立てる際は、分析手法、データの収集方法、収集したデータの公開の3つを中心に検討します。

本フェーズではデータ収集や分析の作業は行いませんが、この時点で想定される分析結果や結論を整理することが重要です。そうすることで、PPDACの全体を俯瞰でき、本フェーズで作成する計画が想定している分析結果や結論に対して適切かを確認することが可能になります。



## 2. 計画の作成

計画の作成では、分析手法・データの収集方法・収集したデータの公開可否の3つを具体化します。

分析手法の検討では、原因仮説の検証に適した手法を選定します。（詳細は、P.42-48を参照）

分析手法を選定したら、必要なデータとその収集方法を計画します。データは様々な形で収集可能ですが、方法によっては時間や費用を要します。プロジェクトの状況や予算に鑑みて、適切な収集方法を検討することが重要です。（詳細は、P.19を参照）

また、委託調査やアンケートなどでデータを収集した場合は、それが公開できるものであるかを検討します。（詳細は、P.31を参照）

## 計画作成における主な注意点

検討事項	検討する際の注意点（例）
分析手法 P.42-48	<ul style="list-style-type: none"> <li>選定した分析手法で仮説は検証できるか</li> <li>担当者のスキルで実施できるか</li> <li>予定しているスケジュールで実施できるか</li> <li>組織が保有するツールで実施できるか</li> <li>外部に委託する場合は予算を確保できるか</li> </ul>
データの収集方法 P.19	<ul style="list-style-type: none"> <li>オープンデータや組織内部に利用できるデータがあるか</li> <li>分析に着手するまでにデータを入手できるか</li> <li>外部にデータ収集を委託する場合やデータを購入する場合は、予算を確保できるか</li> <li>データは電子化された状態で入手できるか</li> <li>収集したデータの格納場所は決まっているか</li> </ul>
データの公開 P.31	<ul style="list-style-type: none"> <li>収集したデータのオープンデータ化は可能か</li> <li>オープンデータ化におけるデータの二次利用も含めた利用許諾はいつ誰からどうやって取得するか</li> </ul>

2 仮説検証のための計画を立てる

**COLUMN** データの収集方法

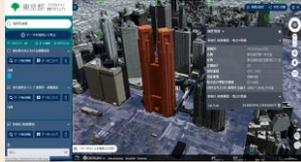
データは様々な方法で収集できます。主な方法としては、オープンデータや組織に蓄積されているデータの収集、事業者からの購入、アンケート調査などがあります。

収集方法によって収集に要する時間や費用が異なるため、プロジェクトのスケジュールや予算を踏まえ、適切な収集方法を検討することが重要です。

例えばオープンデータは国や自治体がデータカタログサイトなどで公開しており、無料かつ短時間で収集できるものです。一方、組織に蓄積されたデータは、課や部を跨いだ共有が問題ないか確認に時間を要することがあります。事業者からデータを購入する場合は費用がかかり、アンケートを実施する場合は費用の他に時間を要することが多いです。

※1：デジタル庁が公開している入力フォームの具体例（巻末付録：リンク一覧②）なども、調査票やアンケートを作成する際の参考にしてください。

代表的なデータの収集方法

収集方法	概要
オープンデータの収集	<p>国や自治体が公開しているデータを収集する。無料かつ短時間で収集しやすい。代表的なサイトとして以下のようなものがある。</p> <p>例) 東京都オープンデータカタログ、e-Govポータル、e-stat、RESASなど</p>
組織に蓄積されているデータの収集	<p>システムに蓄積されているデータや各局が過去に実施し保管している調査結果などを収集する。</p> <p>例) 地理空間に関するデータを収集する際はデジタルツイン庁内データ連携基盤を参照する、など</p> 
事業者からの購入	<p>外部の事業者が収集したデータやそれを元に作成されたレポートを購入する。</p> <p>例) 東京データプラットフォーム上で見つけた事業者から保有データを購入する、など</p>
委託調査、アンケート調査	<p>外部の事業者にデータ収集を委託したり、都が独自にアンケート調査を行う形で収集する。※1</p>

### 3 ワークシートを活用して企画する

## 1. ワークシート作成の背景・目的

本章では、「PPDACワークシート」というツールの概要と使い方を紹介します。

「PPDACワークシート」とは、PPDACの各フェーズを進めるうえで検討すべき項目をまとめたもので、「問題設定シート」と「計画/実行シート」の2つがあります。これらを使用することで、分析に着手する前に各フェーズでの作業内容を俯瞰しながら計画でき、検討漏れや手戻りを予防できます。

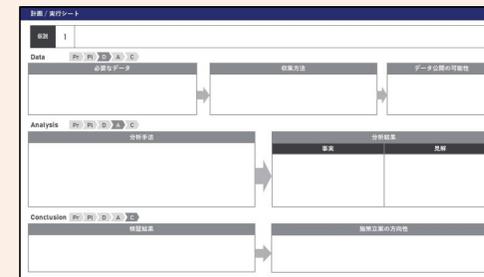
また、これらのワークシートは担当職員同士での認識合わせや上司への報告などにも活用できます。

## PPDACワークシート



### 問題設定シート

あるべき姿と現状のギャップから、データ利活用を通して解消したい問題と検証したい仮説を整理します。



### 計画/実行シート

原因仮説を検証するために必要なデータやその収集方法、分析手法、分析結果、結論を整理します。

### 3 ワークシートを活用して企画する

## 2. ワークシートの構成

PPDACワークシートにはProblem、Data、Analysis、Conclusionの4つのエリアがあります。

### 1 Problemエリア

データ利活用に取り組む背景や、データ利活用を通して解決したい問題とその原因仮説を記入します。

### 2 Dataエリア

仮説の検証に必要なデータや収集方法を記入します。

### 3 Analysisエリア

仮説を検証するための分析手法や分析結果を記入します。

### 4 Conclusionエリア

分析結果を踏まえ仮説が正しかったかどうかの検証結果と、それを踏まえた施策を記入します。

※本ワークシートはPlanフェーズで記入する想定のため、Planエリアはありません

### 問題設定シート

The screenshot shows the 'Problem Setting Sheet' interface. It features a header with 'Problem | Pr | D | A | C'. Below the header, there are three main sections: 'Background' (あるべき姿), 'Problem' (問題), and 'Cause Hypothesis' (原因仮説). The 'Problem' section is highlighted with a large orange '1'. To the left of the 'Problem' section, there is a 'Status' (現状) section. Arrows indicate a flow from 'Background' to 'Problem' and from 'Problem' to 'Cause Hypothesis'. On the right side of the 'Cause Hypothesis' section, there are three numbered boxes (1, 2, 3) indicating a list of hypotheses.

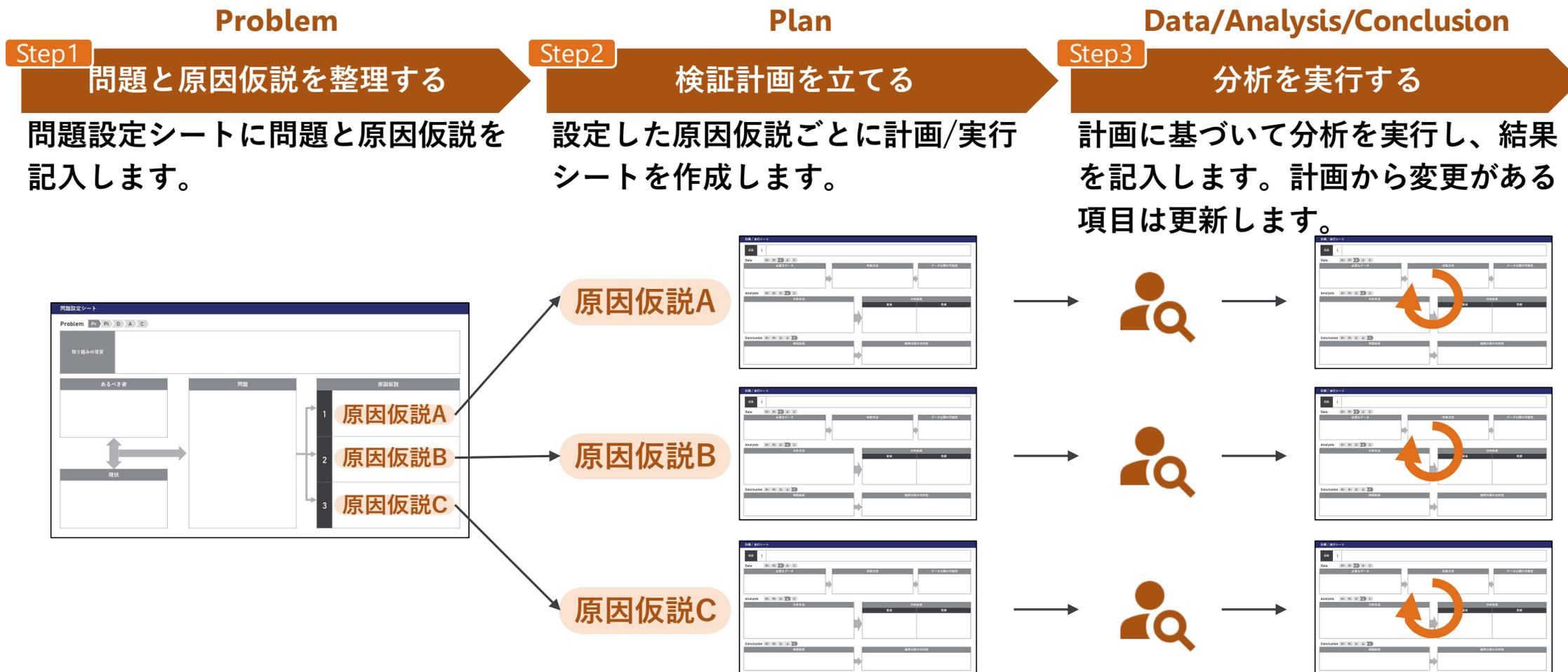
### 計画/実行シート

The screenshot shows the 'Plan/Execution Sheet' interface. It features a header with 'Data | Pr | D | A | C'. Below the header, there are three main sections: 'Data' (Data), 'Analysis' (Analysis), and 'Conclusion' (Conclusion). The 'Data' section is highlighted with a large orange '2' and contains sub-sections for 'Required Data' (必要なデータ), 'Collection Method' (収集方法), and 'Data Collection Possibility' (データ収集の可能性). The 'Analysis' section is highlighted with a large orange '3' and contains sub-sections for 'Analysis Method' (分析手法), 'Fact' (事実), 'Analysis Result' (分析結果), and 'Insight' (見解). The 'Conclusion' section is highlighted with a large orange '4' and contains sub-sections for 'Verification Result' (検証結果) and 'Direction of Policy Formulation' (施策立案の方向性). Arrows indicate a flow from 'Data' to 'Analysis' and from 'Analysis' to 'Conclusion'.

### 3 ワークシートを活用して企画する

## 3. ワークシートの使用方法 全体フロー

ワークシートは以下のステップで記入します。



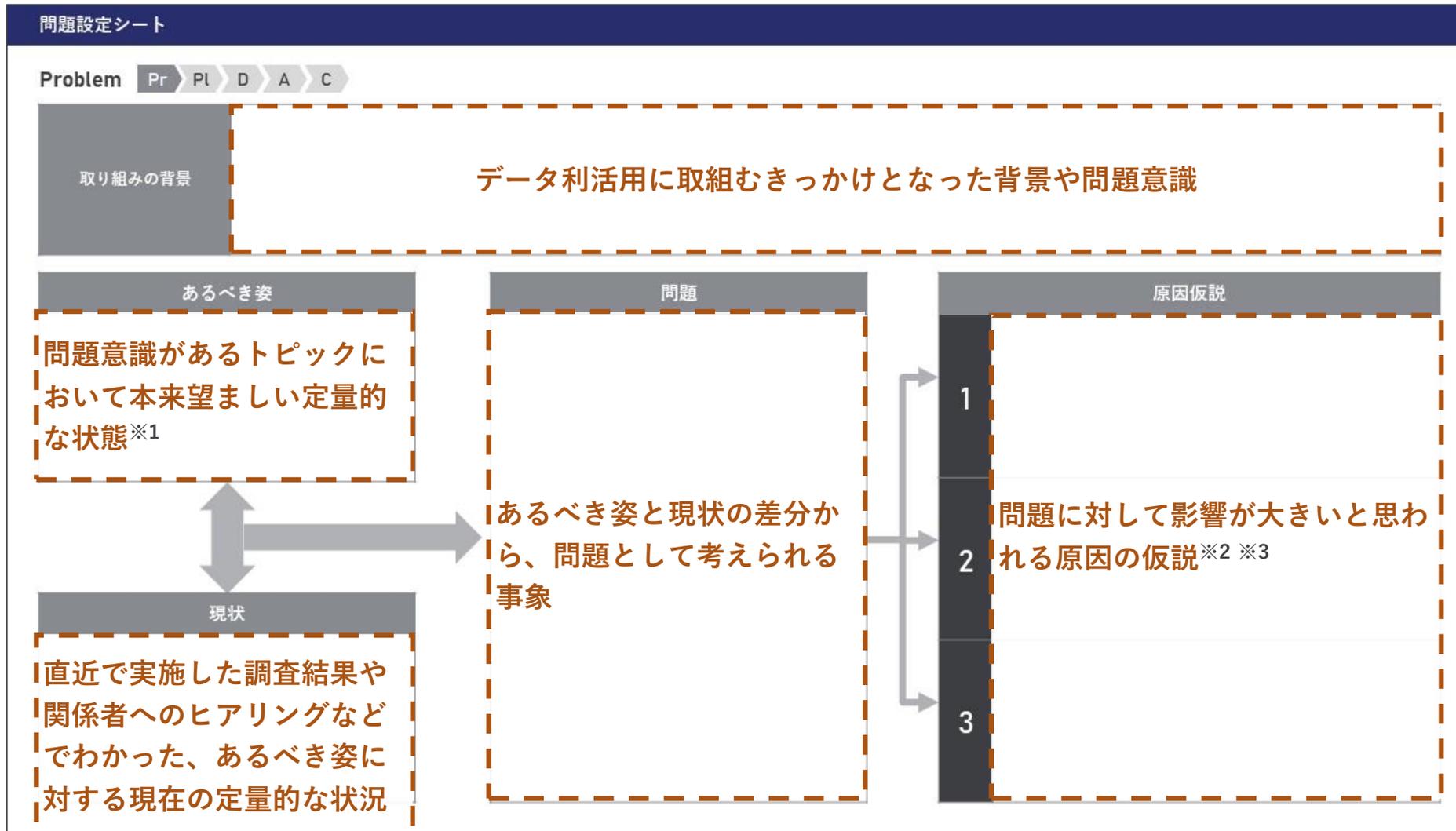
...次頁以降ではページ右上部で各ステップの位置づけを示します

### 3 ワークシートを活用して企画する

## 3. ワークシートの使用方法 Step 1



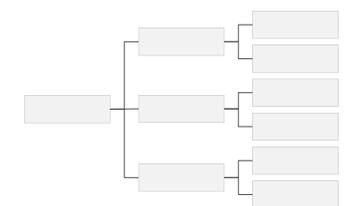
Step1では問題設定シート of 各項目に記入し、分析の前提となる背景・仮説を整理します。



※1：シン・トセイや各  
部局・事業の目標も参考  
にすると良い

※2：原因はこの時点では  
仮説のため、主なものを  
選定した根拠はこの時点  
では必須ではない

※3：下記のようなロ  
ジックツリーで原因を整理  
した上で、影響が大きい  
と思われるものを選定。  
詳しくは、P.15を参照



### 3 ワークシートを活用して企画する

## 3. ワークシートの使用方法 Step 2



Step2では原因仮説それぞれに対して計画/実行シートを記入し、分析結果を予想しながら計画を作成します。

計画 / 実行シート		
仮説	1 問題設定シートに記入した原因仮説	
Data Pr Pl D A C		
必要なデータ	収集方法	データ公開の可能性
仮説を検証するために必要なデータ	必要なデータをどのように収集するか	オープンデータ化するか否か。オープンデータ化しない場合は理由 ※1※2
Analysis Pr Pl D A C		
分析手法	分析結果	
原因仮説をどのような分析手法で検証するか (代表的な分析手法はP.42~48を参照)	事実	見解
	本ステップでは記入不要	分析の結果、指標の大小や増減などの評価として予想されるもの
Conclusion Pr Pl D A C		
検証結果	施策立案の方向性	
分析を踏まえた、仮説が正しかったかどうかの予測。※この時点では原則「正しい」と記入	原因を解消するための施策案	

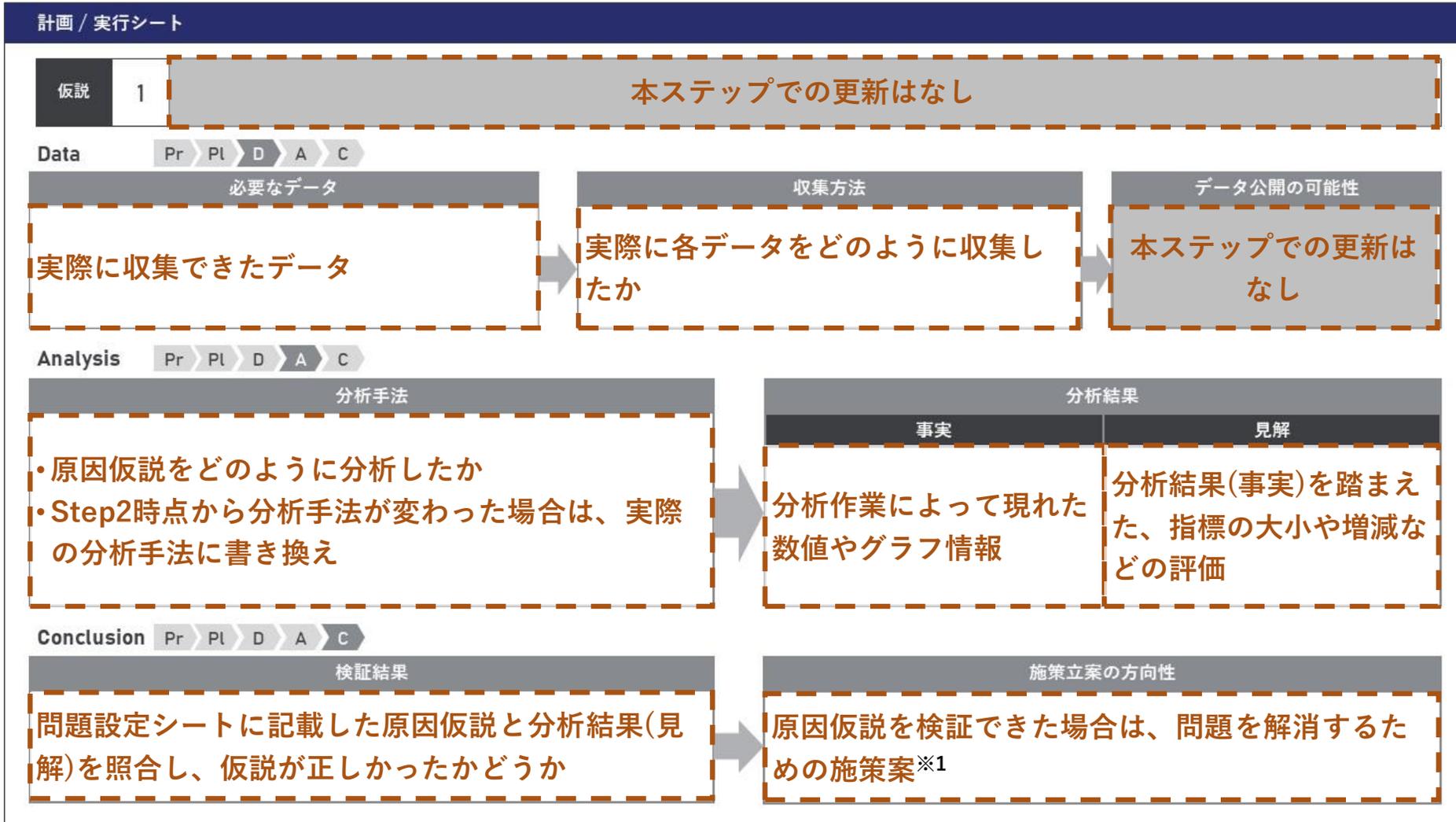
※1：委託調査やアンケート調査などでデータを都が独自に収集する場合のみオープンデータ化の可否を検討

※2：収集したデータは、組織内の他局や組織外の事業者なども活用できるように原則としてオープンデータ化。国および地方公共団体は、オープンデータ化に取り組むことが「官民データ活用推進基本法」で義務付けられている

### 3 ワークシートを活用して企画する

## 3. ワークシートの使用方法 Step 3

Step3では実際に収集したデータや実際の分析結果をもとに、計画時の内容を更新します。



※1：原因仮説を検証できなかった場合は、次に検証する仮説を記入



**Problem** Pr Pl D A C

**取り組みの背景**

- 公共施設Aは郊外に位置し、住民による創作物の展示や販売などを行っている
- 本施設は、年間2億円の売上を上げることを目指しているが目標値を下回っており、来場者数の少なさが問題であると有識者から指摘を受けた。実際に来場者記録データを見てみると特に30歳未満の利用者が少ないため、その原因をデータ分析で特定することにした
- なお、ここでは原因の検証と、原因に対する施策の方向性までを検討範囲としている。より具体的な施策内容は別途検討する予定である

**あるべき姿**

様々な年代層が利用する施設。年代別来場者数の目標は、以下の通り

- 60歳以上：30,000人
- 30~59歳：40,000人
- 30歳未満：30,000人

**現状**

令和5年度の利用者数は、以下の通り

- 60歳以上：33,000人
- 30~59歳：57,000人
- 30歳未満：7,000人

**問題**

30歳未満の利用者が目標に対して23,000人少ない

**原因仮説**

1

駅から距離がありバスの本数も少ないため、車離れの傾向にある若年層にとってはアクセスが悪く来場者が少ない

2

周辺地域の類似した施設と比べて料金が高いため、来場者が少ない

3

施設の広告や宣伝活動があまりされておらず認知されていないため、来場者が少ない

仮説	1	駅から距離がありバスの本数も少ないため、車離れの傾向にある若年層にとってはアクセスが悪く来場者が少ない
----	---	---

**Data**



必要なデータ

本施設および類似公共施設\*1に関する以下のデータ

- ① 最寄り駅から施設までの距離
- ② 最寄り駅から施設までのバスの運行本数
- ③ 30歳未満の1日あたりの平均来場者数

\*1：施設の種類や大きさ、周辺人口が近い施設を「類似公共施設」と定義する

収集方法

データ	収集方法
①	オープンデータカタログサイトから収集
②	インターネット上での調査
③	複数の施設への問い合わせ

データ公開の可能性

- ①②：既に公開情報のため公開は不要
- ③：オープンデータ化する  
(来場者数×日時、年代の情報を取得予定、混雑緩和施策などに活用できる)

**Analysis**



分析手法

**■ 若年層の来場者数とバス本数の関係性を分析**

本施設を含め最寄り駅からの距離が長い施設\*2を対象に、1日のバスの運行本数と30歳未満の1日あたりの平均来場者数を散布図でグラフ化し、バスの運行本数が多いほど来場者数も多い傾向にあることを確認する。

\*2：最寄り駅からの距離が1.6km（徒歩で約20分）以上の施設を「最寄り駅から距離が長い施設」と定義する

分析結果

事実	見解								
<p>1日のバスの運行本数と30歳未満の1日あたりの平均来場者数には以下の関係が見られた。本施設はバスの本数が5本、来場者数が20人</p> <div style="display: flex; align-items: center;"> <table border="1" style="border-collapse: collapse; text-align: center;"> <thead> <tr style="background-color: #eee;"> <th style="font-size: small;">バスの本数</th> <th style="font-size: small;">来場者数</th> </tr> </thead> <tbody> <tr> <td>0本~20本</td> <td>70人以下</td> </tr> <tr> <td>21本~40本</td> <td>40人~90人</td> </tr> <tr> <td>41本~60本</td> <td>80人~90人</td> </tr> </tbody> </table> </div>	バスの本数	来場者数	0本~20本	70人以下	21本~40本	40人~90人	41本~60本	80人~90人	<p>本施設は他施設と比べてバスの本数と来場者が少ない。また、最寄り駅からの距離が長い施設は、バスの運行本数が多いほど来場者も多い傾向にあることから両項目は相関関係にある。</p>
バスの本数	来場者数								
0本~20本	70人以下								
21本~40本	40人~90人								
41本~60本	80人~90人								

**Conclusion**



検証結果

過去に来場者から取得したアンケート結果や有識者との議論もふまえた結果、バス運行本数と来場者数は因果関係にある可能性が高いことがわかった。分析結果から原因仮説は正しいといえる。

施策立案の方向性

若年層の来場者数が多く見込まれる日には最寄り駅から施設までシャトルバスを手配する

# 収集

収集編では、PPDACサイクルにおけるDataフェーズについて説明します。

Dataフェーズは計画通りにデータを収集するだけのフェーズではありません。続くAnalysisフェーズで使用するデータを整え、分析着手前の状態まで持っていくという非常に重要なフェーズです。

本ガイドラインでは「収集」を「データを集めて整備すること」と定義しており、本編ではデータを集める際の考え方や取り扱いだけでなく、データを整備する際の留意事項も説明します。権限や安全性などデータを取り巻く環境も意識しながら、何をすべきか学びましょう。

- |   |               |       |      |
|---|---------------|-------|------|
| 1 | データを収集する      | ————— | P.29 |
| 2 | データ収集における留意事項 | ————— | P.32 |

# ① データを収集する

フェーズ

Pr

Pl

D

A

C

Data

## 1. Dataフェーズとは

Dataフェーズでは、Planフェーズで立てた計画をもとに必要なデータを収集します。後続フェーズでスムーズに分析を進めるためには、このフェーズでデータを集めるだけでなく使える状態に整備する必要があります。データが整備されていないとコンピュータがデータを読み込めない、利活用する権利が無く実際の分析を進められない、といった問題が分析途中に発生してしまいます。

ワークシート更新

Dataエリア

② データの収集における留意事項 で詳しく説明しますがデータを集める際には、利用許諾取得/安全管理、品質管理、標準化という3点を意識することが重要です。

② -1.  
P32~

利用許諾取得/  
安全管理

データを使う権限が自身にあるか？

② -2.  
P35~

品質管理

データの品質は分析で使うに十分か？

② -3.  
P38

標準化

データの定義や表記が揃っているか？

Planフェーズで記載した内容に対し、Dataフェーズでは実際に集めたデータやその方法に書き換えます。



必要なデータ

収集方法

実際に収集できたデータと、  
その際の収集方法に書き換えます。



データ公開の可能性

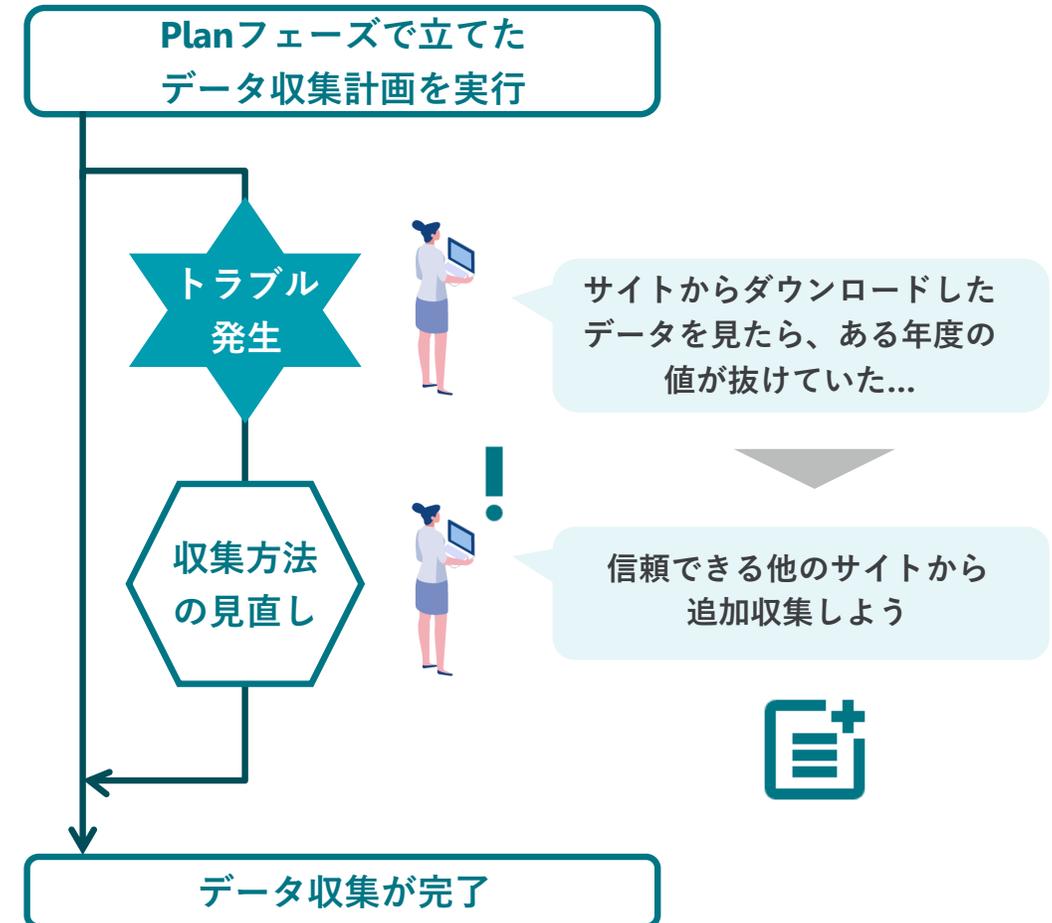
収集したデータをオープンデータ化するか、  
結論を記入します。出来ない場合は理由も  
書きましょう。

## 2. データ収集方法の見直し

データ収集計画はPlanフェーズで作成しますが、データによっては当初予定していた方法では収集できなかったり、データ自体が使えないこともあります。例えば、サイトからダウンロードしたオープンデータの一部が抜けている場合や、アンケート調査の回答数が著しく少ない場合、権限の関係で他部署からデータを共有してもらえない場合などがあります。

そのようなときは、別のサイトから追加収集するなど、他の収集方法を検討してみたり、類似したデータで代用できないか検討しましょう。それでも必要なデータを収集できない場合は、仮説を別の分析手法で検証できないか検討しましょう。

## データ収集方法を見直す流れ



### 3. 収集データの取り扱い

国や地方公共団体が保有する公共データは「官民データ活用推進基本法」で公開が義務付けられています。これはデータ公開によって民間や個人による多様なサービス提供が充実し、都が抱える問題の解決につながることを期待されているためです。

都では行政データをオープンデータ化※1する仕組みとして東京都オープンデータカタログサイトが活用されています。また、データの特性や条件に応じてデータ公開の範囲を柔軟に設定できる東京データプラットフォーム（TDPF）もあります。オープンデータ化が難しい場合は、公開範囲や利用条件を定めることができるシェアードデータとして、東京データプラットフォームに公開することも検討しましょう。

※1：デジタル庁の「オープンデータ基本指針」（巻末付録：リンク一覧③）では、オープンデータとは二次利用可能なルールが適用されており、かつ機械判読に適した、無償で利用できるデータと定義されています。

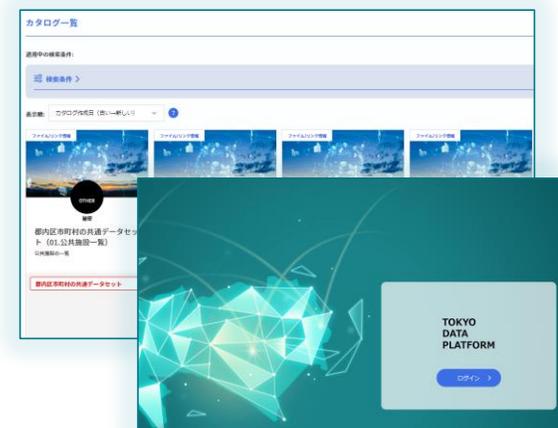
#### 東京都オープンデータカタログサイト

##### ワンポイント

都や都内自治体のオープンデータを検索・利用できるサイトです。こちらにデータを提供することで、利用者はプレビューやダウンロードをして利用できます。



#### 東京データプラットフォーム



##### ワンポイント

行政データだけでなく民間事業者等が保有する様々なデータにつき、検索・利用ができるシステム基盤です。官民の様々なデータを利用できます。（民間データは利用制限あり）

## 1. データの利用許諾取得・安全性管理 利用許諾・安全性管理の必要性

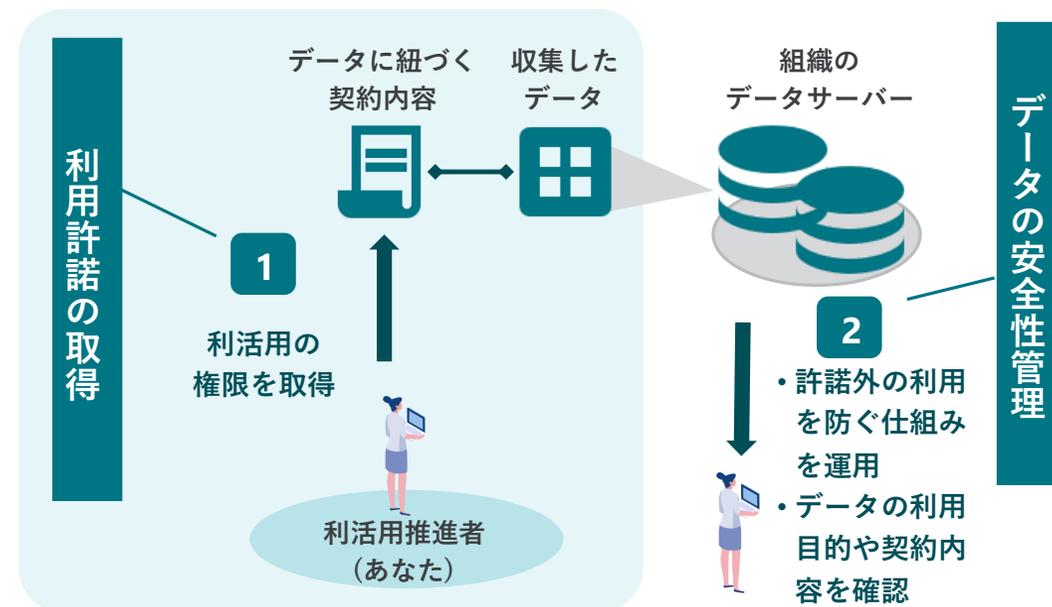
ここでは収集時の留意事項の1点目である「利用許諾取得・安全性管理」について説明します。

Dataフェーズではデータを使用できる状態にする必要がありますが、これはデータの品質としての側面だけでなく、データを利活用しても問題ないという権限としての側面も含まれます。具体的には、特定の目的以外での使用を合意できていない場合やデータの所有権が都にない場合、オープンデータ化の同意を取得できていない場合などがあり、逸脱すると著作権法や個人情報保護法、契約への違反が生じかねません。自分が取得できたから問題ないと考えず、必ず権限も確認しましょう。

また、利用許諾の範囲でデータを扱うことは自分だけが確認すれば良いものではありません。

収集したデータは組織のサーバーに格納され、それを別の誰かが見る可能性もあります。組織としてデータを適切に利活用すべく管理・運用の仕組みを作るといふ、安全性管理の観点も非常に重要です。

## データ利活用における望ましい状態



# 1. データの利用許諾取得・安全性管理 利用許諾における注意事項

データには様々な流れがあり、各々の利用場面に応じてデータ利用許諾の注意事項も異なります。

収集や利活用が組織内で完結する場合は、アンケートの回答者などデータの提供者から利用目的や利用範囲の許諾を得ることが必要です。調査委託や購入など組織外からデータ収集する場合は、1つの事業や部署に閉じた限定的な利用目的・範囲の合意だけでなく、オープンデータ化の許諾やローデータ※1の受取り、受領データの継続的な利用権限などについても確認することが重要です。

データを組織外に提供する場合、使用したデータの削除や第三者への提供禁止などを事前に合意してください。

利用許諾に関するより詳細な内容は、経済産業省の「データ利活用のポイント集」を参照ください。

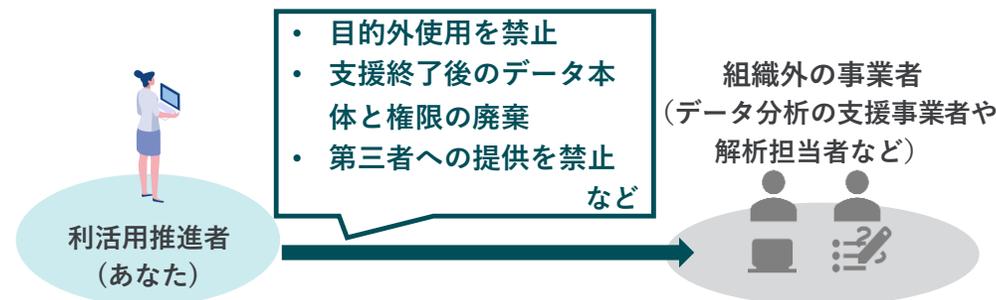
※1：ローデータとは、不要なデータの削除や集計などの加工を行う前のデータを意味します。例えば、アンケート分析を行う場合、分析を行う前の生の回答データがローデータです。

## 1 利用許諾の取得

### データを収集し利活用するとき



### データを外部へ提供するとき



## 1. データの利用許諾取得・安全性管理 安全性管理における注意事項・考え方

安全性管理は、仕組みとルールの両面から考えます。

ここでの仕組みとは、許諾の範囲を超えた活用や、業務上のミスによる契約違反などが発生しないようにするための取組みです。具体的には、機密性の高いデータや個人情報を含むデータには厳しいアクセス権限を設定するよう管理者に依頼したり、組織内で収集したデータと外部から収集したデータのように許諾範囲が異なるデータは異なるフォルダに格納する、などの取組みが必要です。

しかし、仕組みだけで管理するのは現実的ではありません。データを収集・利活用・提供するときには、各職員が必ずデータの利用目的や利用範囲などの権限を確認することが重要です。

## 2 データの安全性管理

許諾外の利用を防ぐ仕組み

## アクセス制限

- ・ 許諾範囲で利用する職員にのみデータへのアクセス権限付与するよう管理者に依頼

## 機密情報管理

- ・ 組織内で収集したデータと外部から収集したデータは異なるフォルダに格納する
- ・ データのマスキング
- ・ 個人情報の除外

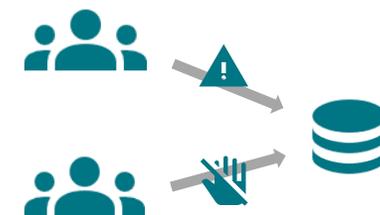
## パスワード管理

- ・ データアクセスへのパスワード設定と定期的な変更

許諾の通りに利活用するルール

## 権限の確認

- ・ データ利用時に、利用目的や契約内容を確認



自分のデータ利活用後も、組織内のデータが安全に契約の範囲内で用いられる状況を管理します

## 2 データ収集における留意事項

## 2. データの品質管理 品質管理の必要性

収集時の留意事項の2点目は「データの品質管理」です。

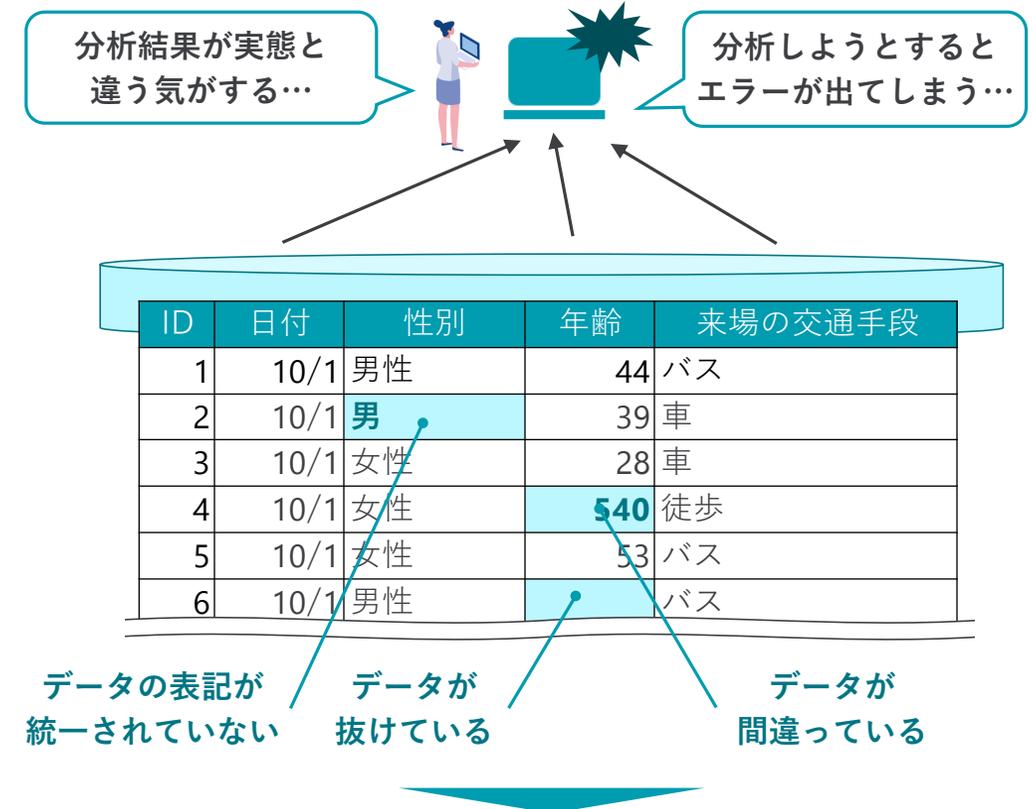
データの品質を適切に管理することはサービスの品質を高めるだけでなく、個々の作業負荷の軽減につながるなど、様々なメリットがあります。

元のデータの品質が低いと、それを加工したデータの品質もその影響を受け、結果としてデータを利用したサービス自体の品質も下げてしまいます。また、データの利用者也データ品質の改善作業に取り組むこととなり、多くの時間が費やされてしまいます。

本節ではデジタル庁の「データ品質管理ガイドブック」

※1に基づきデータの品質についての考え方を示します。

## データ品質の必要性



正しい分析結果を得るためには  
データの品質が重要

※1：データ品質に関するより詳細な情報を確認したい場合は、デジタル庁の「データ品質管理ガイドブック」（巻末付録：リンク一覧④）を参照してください。

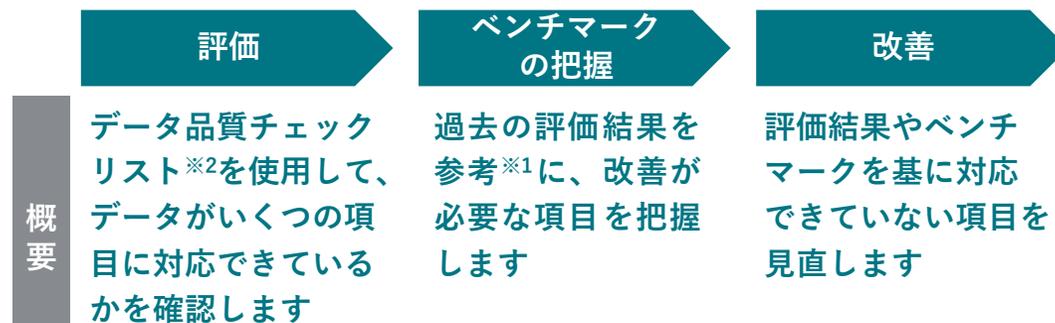
## 2. データの品質管理 データ品質の評価

データを収集する際は、それが高い品質のデータであるかを客観的に評価して改善することが重要です。過去に類似したデータを評価している場合は、右図のようにその評価結果をベンチマーク※1として活用し改善します。ベンチマークを把握することによって、評価結果が良いかどうかを客観的に判断できるようになるためです。

なお、実際の業務において調査票やアンケートなどで新たにデータを生成する場合は、チェックリスト編で説明する「データ品質チェックリスト」※2を使用して右図のようにデータ品質を評価・改善します。

次頁では、データ品質を評価するにあたっての重要な観点を説明します。

### データ品質の改善プロセス



概要

データ品質チェックリスト※2を使用して、データがいくつの項目に対応できているかを確認します

過去の評価結果を参考※1に、改善が必要な項目を把握します

評価結果やベンチマークを基に対応できていない項目を見直します

例

作成した調査票をチェックリストで評価したところ、8つの項目をクリアした

過去に作成した調査票は、10個の項目をクリアしていた

過去にクリアしていた項目をクリアできるように、アンケートの設計を見直す

評価結果

No	チェック項目	
1	入力欄の書式を指定しているか	✓
2	入力できる内容をプルダウン形式などで指定しているか	✓
3	明らかな誤入力を防ぐ設計にしているか	

過去の評価結果

No	チェック項目	
1	入力欄の書式を指定しているか	✓
2	入力できる内容をプルダウン形式などで指定しているか	✓
3	明らかな誤入力を防ぐ設計にしているか	✓

No.3をクリアできるように、年齢を入力する欄は3桁以上の数字を入力できないようにしましょう



※1：ベンチマークとは、他の組織や過去の評価結果など、比較のために使用する指標を指します。過去の評価結果など対象がある場合はベンチマークの把握を実施しましょう  
 ※2：チェックリストの利用シーンや使用方法などの詳細は本ガイドラインのチェックリスト編「データ品質チェックリスト」を参照ください

## 2 データ収集における留意事項

## 2. データの品質管理 データ品質の観点

データの品質※1は、正確性、完全性、一貫性、可用性、最新性の5つの観点で評価することが重要です。それぞれの概要は以下に示す通りです。

- 正確性：誤字脱字などがなく、データが正しいこと
- 完全性：データに空欄がなく、網羅的であること
- 一貫性：データ内の項目に矛盾がないこと
- 可用性：データが利用可能な状態にあること
- 最新性：データが最新の状態に維持されていること

上記5つの観点に関する具体的なチェック項目やチェックリストの利用場面についてはチェックリスト編の「データ品質チェックリスト」を参照ください。

## データ品質の観点（具体例）

	良い例	悪い例												
正確性	<table border="1"> <thead> <tr> <th>性別</th> <th>年齢</th> </tr> </thead> <tbody> <tr> <td>女性</td> <td>28</td> </tr> <tr> <td>女性</td> <td>54</td> </tr> </tbody> </table>	性別	年齢	女性	28	女性	54	<table border="1"> <thead> <tr> <th>性別</th> <th>年齢</th> </tr> </thead> <tbody> <tr> <td>女性</td> <td>28</td> </tr> <tr> <td>女性</td> <td>540</td> </tr> </tbody> </table> <p>誤字がある</p>	性別	年齢	女性	28	女性	540
性別	年齢													
女性	28													
女性	54													
性別	年齢													
女性	28													
女性	540													
完全性	<table border="1"> <thead> <tr> <th>性別</th> <th>年齢</th> </tr> </thead> <tbody> <tr> <td>女性</td> <td>53</td> </tr> <tr> <td>男性</td> <td>21</td> </tr> </tbody> </table>	性別	年齢	女性	53	男性	21	<table border="1"> <thead> <tr> <th>性別</th> <th>年齢</th> </tr> </thead> <tbody> <tr> <td>女性</td> <td>53</td> </tr> <tr> <td>男性</td> <td></td> </tr> </tbody> </table> <p>空欄がある</p>	性別	年齢	女性	53	男性	
性別	年齢													
女性	53													
男性	21													
性別	年齢													
女性	53													
男性														
一貫性	<table border="1"> <thead> <tr> <th>性別</th> <th>年齢</th> </tr> </thead> <tbody> <tr> <td>男性</td> <td>44</td> </tr> <tr> <td>男性</td> <td>39</td> </tr> </tbody> </table>	性別	年齢	男性	44	男性	39	<table border="1"> <thead> <tr> <th>性別</th> <th>年齢</th> </tr> </thead> <tbody> <tr> <td>男性</td> <td>44</td> </tr> <tr> <td>男</td> <td>39</td> </tr> </tbody> </table> <p>表記が統一されていない</p>	性別	年齢	男性	44	男	39
性別	年齢													
男性	44													
男性	39													
性別	年齢													
男性	44													
男	39													
可用性	<table border="1"> <thead> <tr> <th>名称</th> <th>更新日時</th> </tr> </thead> <tbody> <tr> <td>アンケート結果.xlsx</td> <td>2023/10/30</td> </tr> </tbody> </table>	名称	更新日時	アンケート結果.xlsx	2023/10/30	<table border="1"> <thead> <tr> <th>名称</th> <th>更新日時</th> </tr> </thead> <tbody> <tr> <td colspan="2">(データなし)</td> </tr> </tbody> </table> <p>ファイルが壊れている</p>	名称	更新日時	(データなし)					
名称	更新日時													
アンケート結果.xlsx	2023/10/30													
名称	更新日時													
(データなし)														
最新性	<table border="1"> <thead> <tr> <th>名称</th> <th>更新日時</th> </tr> </thead> <tbody> <tr> <td>アンケート結果.xlsx</td> <td>2023/10/30</td> </tr> </tbody> </table>	名称	更新日時	アンケート結果.xlsx	2023/10/30	<table border="1"> <thead> <tr> <th>名称</th> <th>更新日時</th> </tr> </thead> <tbody> <tr> <td>アンケート結果.xlsx</td> <td>2023/10/6</td> </tr> </tbody> </table> <p>データが古い</p>	名称	更新日時	アンケート結果.xlsx	2023/10/6				
名称	更新日時													
アンケート結果.xlsx	2023/10/30													
名称	更新日時													
アンケート結果.xlsx	2023/10/6													

※1：デジタル庁の「データ品質管理ガイドブック」（巻末付録：リンク一覧④）では、データ品質に関して15個の評価項目を示しています。都では、特にデータの中身に関する項目として上記の5項目を重点項目として位置付けています。

**COLUMN** 必要なデータ品質の水準

データの品質は高ければ高いほど良いとは限りません。なぜなら、品質を高めるには費用や手間を要するからです。そのため、データを収集する際は利用目的に照らし、必要な品質の水準を確認することが重要です。

例えば、医療現場で患者に投与する薬剤のなかには、データの精度が非常に細かい単位まで求められるものもあります。そうした薬剤は、投与量に誤りがあると人命にも関わる可能性もあるため、1日に投与できる量の上限が細かい単位まで定められています。

一方、一般家庭で熱があるかを判断する際には平熱なのかが分かれば良いため、体温計では小数点1桁以下の単位まで高い精度のデータは必要ありません。

医療現場での投薬



日付	投与量(単位略)
10/1	80.0
10/2	80.0
10/3	70.0

高い正確性が必要

一般家庭での体温測定



体温を測ったら36度台だったから問題なさそうだ

**ワンポイント**

データ品質において最新性は重要な1観点ですが、高頻度なデータ更新には手間や費用を要します。必ずしも全データの更新が高頻度である必要はなく、利用目的を踏まえて適切な更新頻度を考えましょう。



低頻度

クレジットカードの利用額は最終的な金額さえ分かればよいため、日次の更新でもよい



高頻度

銀行口座の残高は金額にずれがあると取引に支障をきたすため、リアルタイムでの更新が必要

## 2 データ収集における留意事項

### 3. データの標準化 標準化の必要性

収集時の留意事項の3点目は「標準化」です。

標準化とは、データの項目名や表記形式、定義を揃えることで、デジタル庁では、拡張性が高く、連携が容易なデータを設計することで、データの利活用をスムーズに行えるように「政府相互運用性フレームワーク（GIF）」を提供しています。

例えば、自治体への公開ニーズが高いデータの標準化にGIFを適用して、「自治体標準オープンデータセット」を規定しています。

データを収集し利活用する際は、標準化した規格があるか事前に調査し、「自治体標準オープンデータセット」のような規格がある場合は、活用するようにしましょう。

また、データの標準化に関連する重要な項目の1つに、ベース・レジストリ※1があります。ベース・レジストリは、国全体のデジタル化を進めるためのデジタルインフラで、行政機関間の情報連携や民間事業者を含めたデータの利活用を推進するためのデータベースです。システム構築などデータベースを設計する際は、ベース・レジストリとの連携を見据え、国の方針や仕様と足並みを揃えることも重要です。

#### ベース・レジストリのイメージ



※1：内閣官房情報通信技術（IT）総合戦略室の「ベース・レジストリの指定について」（巻末付録：リンク一覧⑤）では、ベース・レジストリとは「公的機関等で登録・公開され、様々な場面で参照される、人、法人、土地、建物、資格等の社会の基本データであり、正確性や最新性が確保された社会の基盤となるデータベース」と定義されています。

# 利活用

利活用編では、PPDACサイクルのAnalysisフェーズとConclusionフェーズについて順に説明します。

「分析」と称される作業の内容は、その目的や原因仮説、データやツールにより異なり、ケースバイケースです。本編ではどのような分析観点や特徴があるのか紹介しますので、皆さんが考える課題を検証するにはどんな分析を行うのが良いか、作業内容をイメージしながら読み進めましょう。

また最後には、Analysis/Conclusionフェーズをまたいだ具体的な実践例も紹介しますので、あわせてお読みください。

- |   |           |       |      |
|---|-----------|-------|------|
| ① | データを分析する  | ————— | P.41 |
| ② | 分析結果を考察する | ————— | P.53 |
| ③ | 利活用の実践例   | ————— | P.56 |

# ① データを分析する

## 1. Analysisフェーズとは

Planフェーズでは仮説の検証方法を計画し、Dataフェーズでは実際に使用するデータを収集しました。Analysisフェーズではこれまで立てた計画にもとづき実際に分析作業を実施します。この作業結果が、仮説の正しさを最終的に判断するConclusionフェーズに繋がります。仮説の正しさを検証するには、検証したい仮説に対して正しい分析観点・手法を選択する必要があります。

これから紹介する主な分析観点の特徴や活用シーン、注意点を学び、選ぶべき分析観点や表現手法をイメージできるようにしましょう。



### ワークシート更新

### Analysisエリア

Planフェーズでは想定内容を記載しましたが、Analysisフェーズではこの記述を実際の分析を通して得られた結果に書き換えます。



#### 分析結果（事実）

分析から得られたグラフの情報や計算結果の値など、データから読み取れる数値情報を記入

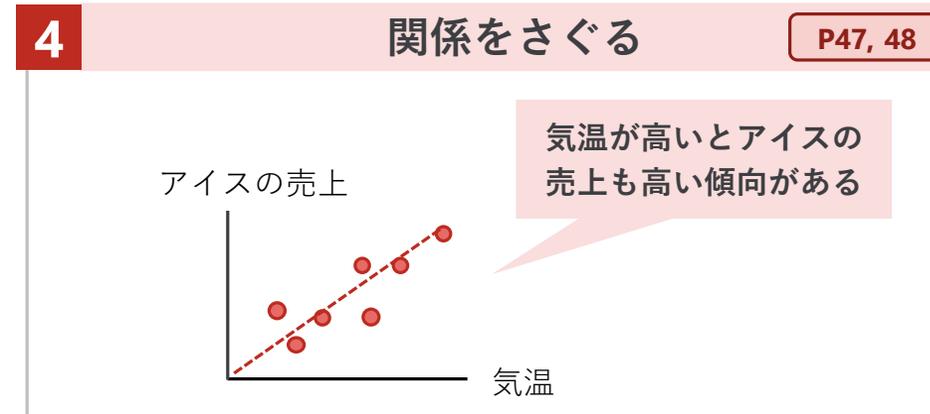
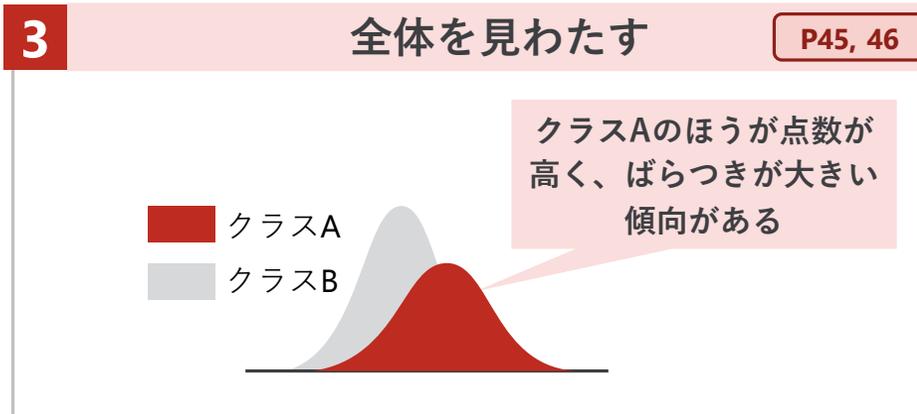
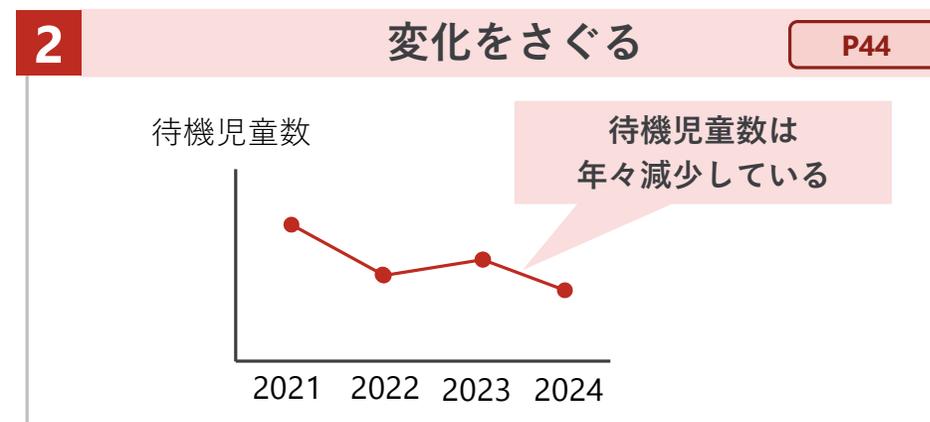
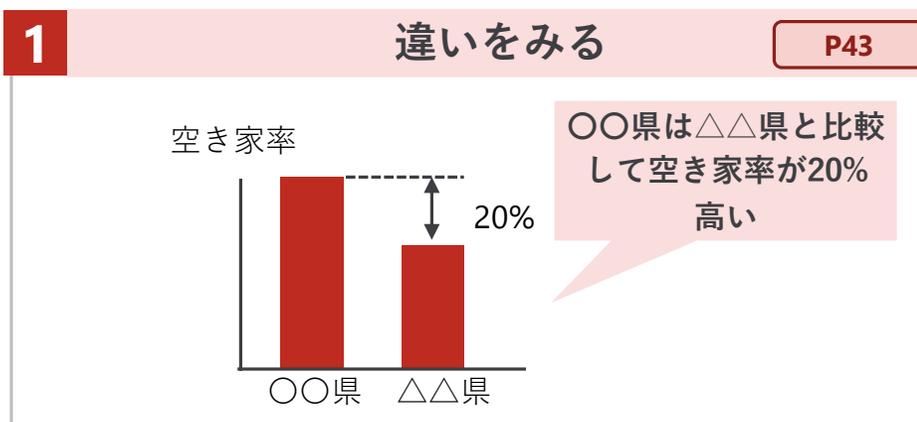


#### 分析結果（見解）

「分析結果（事実）」からどのようなことが言えるか、指標の大小や増減などの評価を記入

## 2. データを分析する観点

よくあるデータ分析方法として4つの観点があげられます。  
仮説や課題に応じて、これらの観点を上手に使い分けましょう。



1 データを分析する

2. データを分析する観点①：違いをみる

どんな  
観点？

実数や比率の差から示唆を抽出

主な利用  
シーン

1つの指標に関して複数項目間の状況や傾向の違いを把握したいときに使用（施策実施の前後比較や数字の大小比較など）

代表的な  
グラフ

棒グラフや円グラフ

具体例

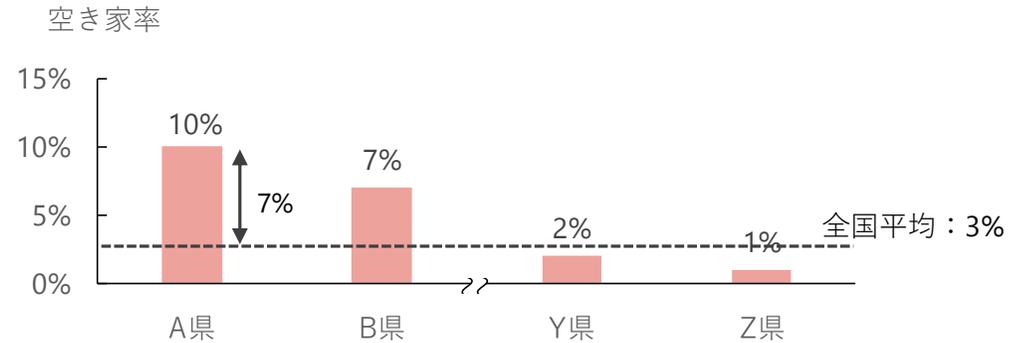
例1 都道府県別の空き家率の比較

都道府県別の空き家率を示している。A県は都道府県のなかで最も空き家率が高いため、空き家が発生しやすい要因がある

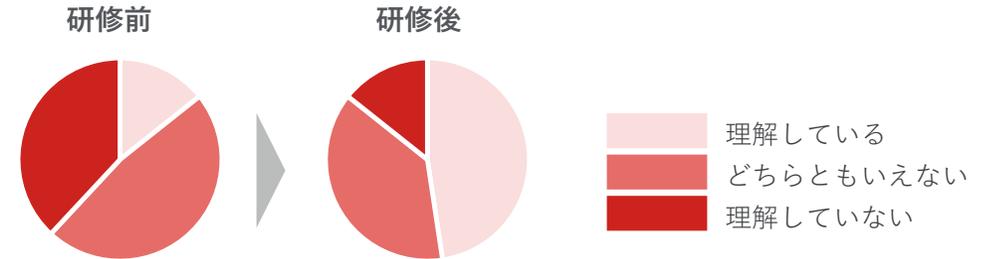
例2 研修前後の理解度の比較

研修前後の理解度の比較を示している。研修後は「理解している」の割合が増えたことから、研修を通して理解度が上がったことがわかる

例1 都道府県別の空き家率の比較



例2 研修前後の理解度の比較



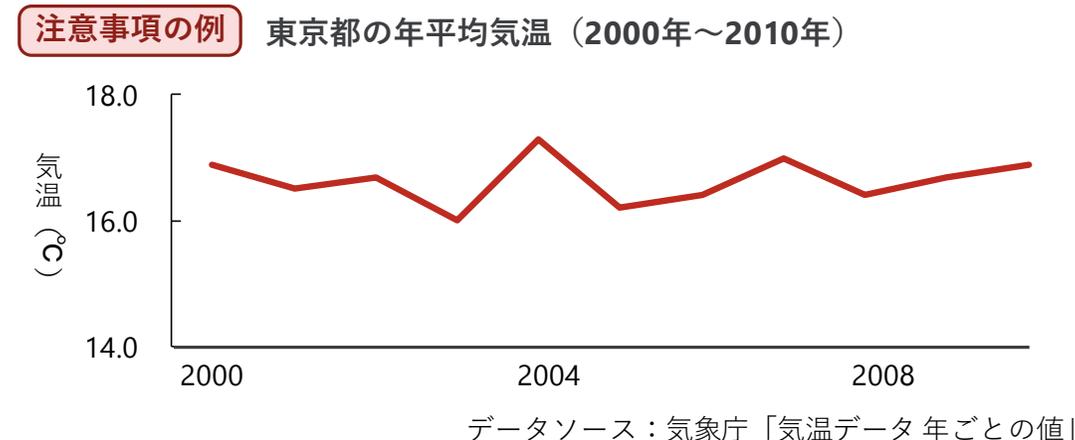
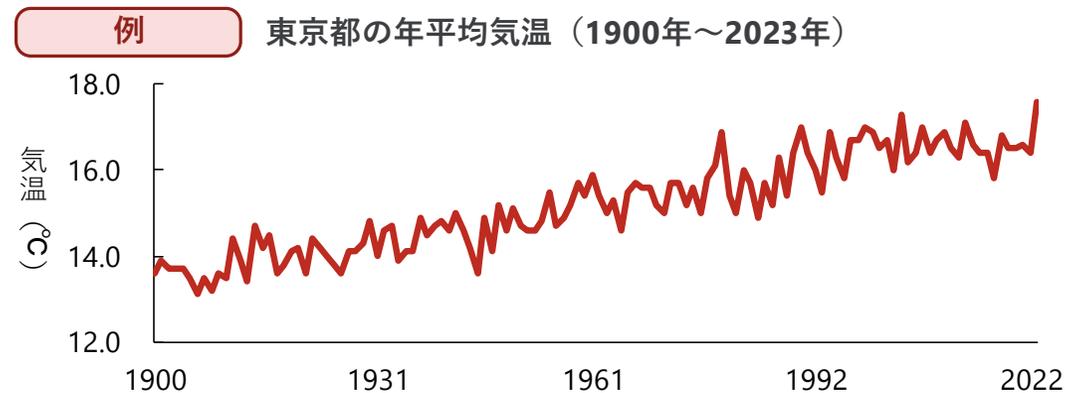
ワンポイント

施策の効果を測るときは、施策実施前のデータ収集を忘れずに実施しましょう

# 1 データを分析する

## 2. データを分析する観点②：変化をさぐる

どんな観点？	時間軸の推移から示唆を抽出
主な利用シーン	一定の集計期間にわたって、指標の傾向や状況を把握したいときに使用（数値の変動傾向の把握など）
代表的なグラフ	折れ線グラフ
具体例	<p><b>例</b> 東京都の年平均気温（1900年～2023年）</p> <p>1900年から2023年までの東京都の年平均気温を示している。多少の増減はあるものの上昇し続けているため温暖化が進んでいる</p>
注意事項	<p><b>注意事項の例</b> 東京都の年平均気温（2000年～2010年）</p> <p>データの集計期間によっては、グラフの傾向が変わり得られる示唆も変わってしまうことに注意。右図「注意事項の例」では、集計期間が10年のみのため、気温が変化していない</p>



### ワンポイント

集計期間によってグラフの傾向も変わるためデータの特性や分析の目的に合わせて、適切な集計期間を設定しよう

## 2. データを分析する観点③：全体を見わたす

どんな  
観点？

データの広がり具合などから示唆を抽出

主な利用  
シーン

1つの指標に関して、どの値にどの程度データが集中しているのかを把握したいとき（数字の分布状況の調査や異常値の発見など）

代表的な  
グラフ

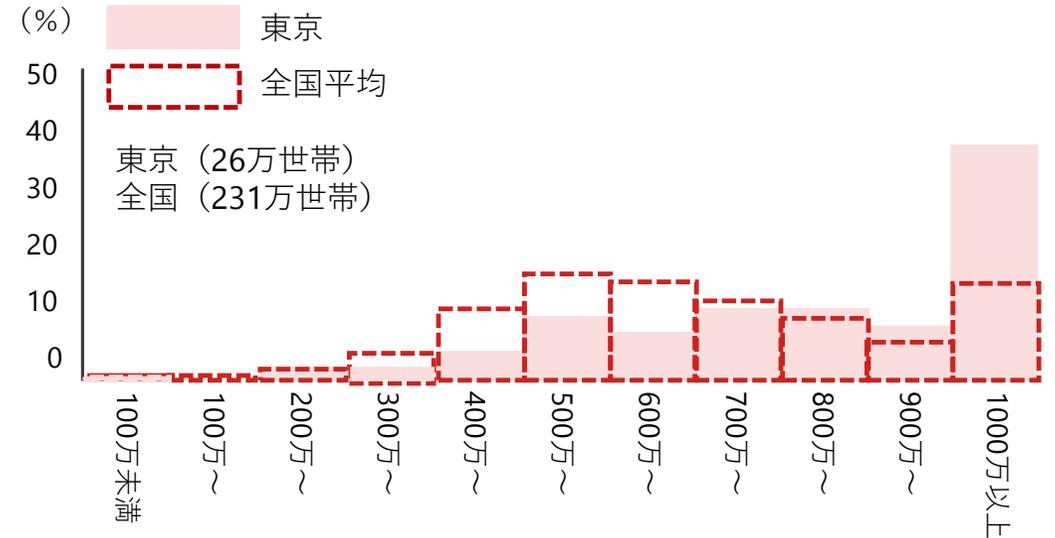
ヒストグラム※1

具体例

例 30代子育て世代の世帯年収分布図

東京都の子育て世代の年収分布を表している。東京都のデータが全国平均よりも右側に集中していることから、東京都のほうが年収の高い世帯が多いことがわかる

例 30代子育て世代の世帯年収分布図



データソース：総務省「就業構造基本調査」

ワンポイント

データがどこに集中しているか、より定量的に把握するためには次ページ紹介の統計的な数値を取り入れると良い

※1：ヒストグラムとは横軸にデータのカテゴリーを、縦軸にそのカテゴリーに含まれるデータの個数を示したグラフです

2. データを分析する観点③：全体を見わたす 発展編

「全体を見わたす」では、グラフの他に代表値や散布度からも示唆を得られます。

代表値

代表値とは、データの分布状況を1つの値で代表した数値です。データの分布傾向を客観的に素早く把握したい際に使用されます。主な代表値とその定義は以下のとおりです。

生徒	A	B	C	D	E
点数	20	34	38	44	44

平均値	全ての値を足してサンプル数で割った値	36
中央値	順番に並び替えた際に真ん中にくる値	38
最頻値	最も頻繁に表れる値	44

上記のうち、特に使い分けに注意が必要なのが**平均値**と**中央値**です。

**極端な値（外れ値）**が多く含まれる場合は、中央値のほうが偏りなくデータの状況を捉えられるため**中央値**を活用しましょう。

例えば、上記の例に100点を取ったFさんが加わると平均値は46.7になり大きく上振れてしまいます。一方で中央値は41<sup>\*1</sup>のままであり、クラスの実態をより正確に把握できます。

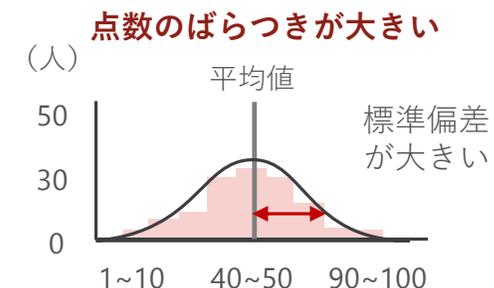
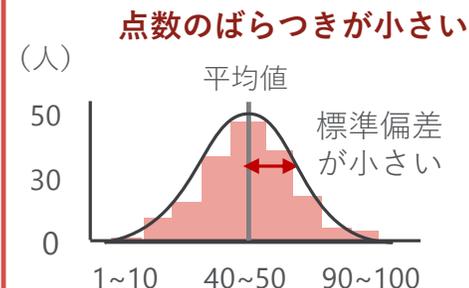
※1：データの数か偶数の場合は、「真ん中の値」が2つ登場するため、その2種類を足して2で割ったものを中央値とします

散布度

散布度とは、データのばらつき度合を示す指標です。データの散らばり度合を客観的に素早く把握したい際に使用されます。主な散布度は標準偏差で、その定義は以下のとおりです。

生徒	A	B	C	D	E	
点数	20	34	38	44	44	44点から平均値の36点を引く
偏差	-16	-2	2	8	8	

偏差	各値－平均値	表通り
標準偏差	各値がどの程度平均値から離れているかを示す	8.9



はじめに

企画

収集

利活用

チェックリスト

1 データを分析する

2. データを分析する観点④：関係をさぐる

どんな  
観点？

2つの変数の関連性から示唆を抽出

主な利用  
シーン

2つの指標に関して、一方の指標が変化した場合にもう一方の指標はどのように変化する傾向にあるか調べたいとき（要因の調査など）

主な  
グラフ

散布図

具体例

例 気温とアイスの売上の散布図

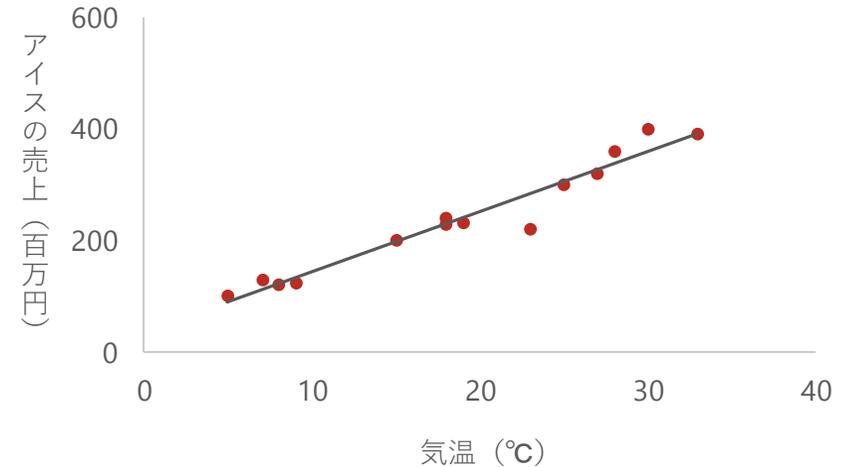
気温とアイスの売上の分布を示している。直線状に左下から右上へ延びているため気温が高いとアイスの売上も高い傾向にあることがわかる

ワンポイント

2つの変数の関連性を相関関係と呼称し、その指標として「相関係数」がある。補足のとおり、相関係数は-1から+1の範囲を取り、0であれば相関関係がなく、±1に近いほど強い相関関係があることを示す。

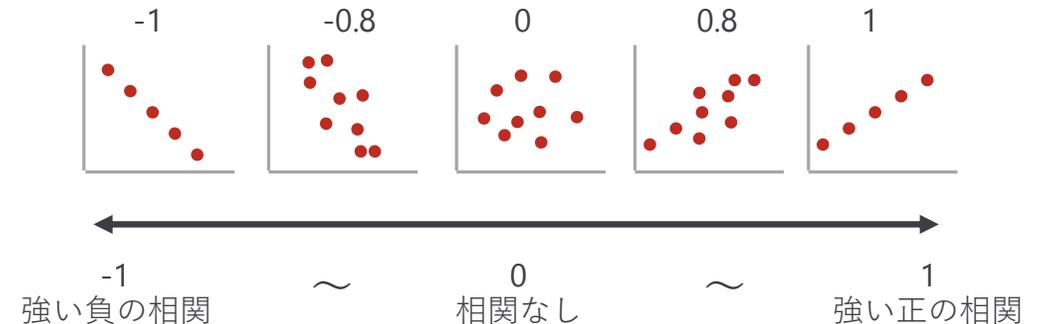
例

気温とアイスの売上の散布図



補足

相関係数の範囲・散布図との関連性



## ① データを分析する

## 2. データを分析する観点④：関係をさぐる 発展編

相関関係のある2つの変数には因果関係があるものもあります。因果関係とは、相関関係のなかでもAを原因としてBが変動する関係にあることを指します。

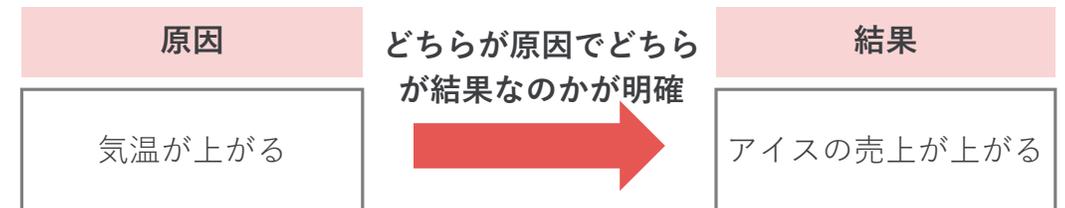
例えば、「気温とアイスの売上」は、右図の①のとおり原因と結果に分類できるため相関関係のみならず因果関係にもあると言えます。

一方で、相関関係ではあるものの因果関係にないものも多くあります。例えば、「特定地域におけるアイスの消費量とアイス販売する店舗の数」に相関があったとしても、右図の②のとおり明確にどちらが原因でどちらが結果かを判断できないため因果関係にはありません。相関関係と因果関係を混同しないようにしましょう。

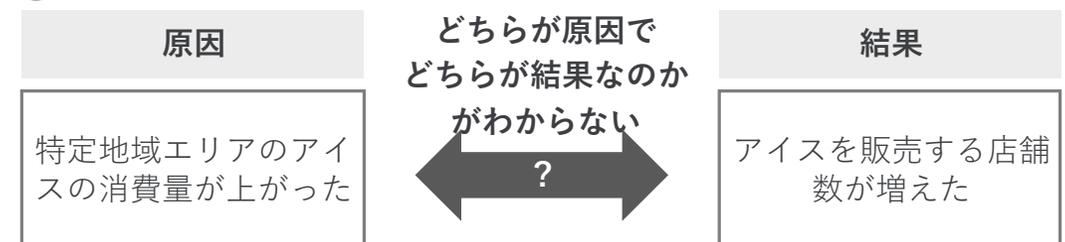
## 相関関係と因果関係



## ① 相関関係かつ因果関係にある場合



## ② 相関関係にあるが因果関係にない場合



## 3. 分析実施時の注意事項①：同じ条件で比較する

これまで述べてきた4つの分析観点を正しく使い分けていたとしても、分析に使用する数値の捉え方が間違っていると誤った判断につながる可能性があります。

よくある間違いとして多いのは、2つの要素を異なる条件で比較することです。例えば、A市とB市における新型コロナウイルスの感染者数を比較する際に、百人あたりの感染者数ではなく、感染者数をそのまま比較してしまうことが挙げられます（図1）。また、「若者」のみの感染者数を比較する場合、百人あたりで比較したとしても「若者」の定義が異なっていると誤った結果を導き出してしまいます（図2）。要素を比較する際は、人口などの全体の数も考慮して比較できているか、比較対象は本当に同じものなのか、考えてみましょう。

図1 新型コロナウイルス感染者数の比較

	感染者数	人口（人）	100人あたりの感染者数※1
A市	300	1,200	25
B市	700	10,000	7

※1：百人あたりの感染者数=感染者数/人口×100

実数で見るとB市のほうが感染者をより多く出しているように見える

しかし、百人あたりの感染者数だとA市の方が多いことがわかる

図2 百人あたりの感染者数（若者）の比較

	A市	B市
若者の定義	29歳以下	39歳以下
100人あたりの感染者数（若者）	120	160

A市とB市で若者の定義が異なるため、比較するためには定義を合わせる必要がある

## ① データを分析する

### 3. 分析実施時の注意事項②：サンプルの考え方

データ分析を実施する際は前述した注意事項以外にも、**サンプル※1**が少なかったり収集対象に偏りがあると信頼できる分析結果が得られなくなることに注意しましょう。

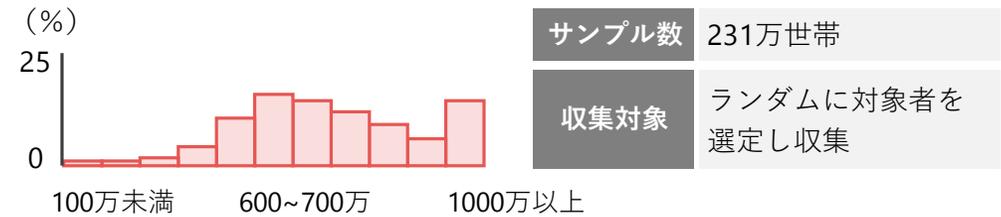
図1と図2はともに世帯年収のヒストグラムですが、結果に大きく差が生じています。これは、サンプル数や収集対象に差があったためです。信頼性の高いデータを得るためには、図1のように一定のサンプル数を確保したり、幅広い人に回答いただくための取組みが必要です。

実際に分析をする際にも、サンプル数は十分か、結果に偏りが出るような収集方法でないか、他のレポートを確認したり周りの人に相談しながら考えましょう。

※1：サンプルとは、統計調査において母集団から抽出したデータを指し、「標本」とも言います。

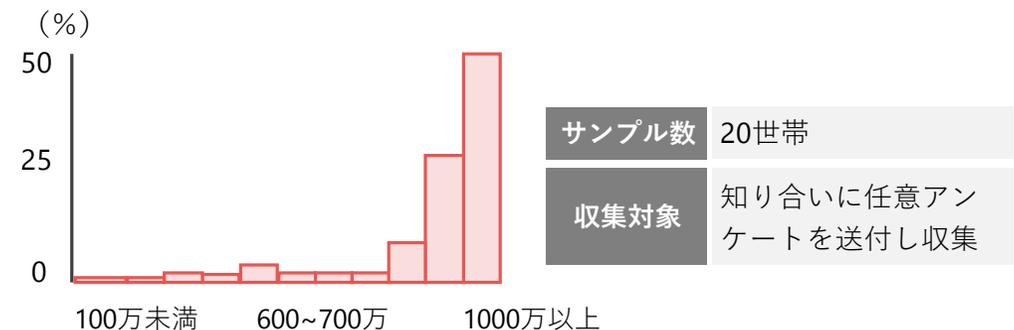
#### 子育て世代の世帯年収（全国平均）

✓ **図1** サンプル数が十分であり収集対象に偏りが無い場合



データソース：就業構造基本調査

✗ **図2** サンプル数が少なく、収集対象に偏りがある場合



1 データを分析する

3. 分析実施時の注意事項③：外れ値の取り扱い

データ分析実施時は、外れ値の取り扱いにも気を付ける必要があります。外れ値とは、他の値と比べて極端に小さな値、あるいは極端に大きな値のことです。

データに外れ値が存在していると、分析結果が大きく歪められてしまうことがあります。例えば、図1は、外れ値がないと相関係数※1は0.89ですが、外れ値があると0.53になってしまいます。

データに外れ値がある場合は、なぜ外れ値が存在するのかデータの提供元に確認するなど調査を行い除去すべきかを決めましょう。例えば、入力ミスなどの誤りで発生した外れ値は修正するか、取り除くべきです。一方で、誤りでない外れ値に関しては、図2のように外れ値も含めて分析したほうが良いか考えて判断しましょう。

※1：相関係数はP.47の「関係をさぐる」にて解説

B市図書館の利用者層と本貸出回数

図1 外れ値なし・ありの場合の相関係数

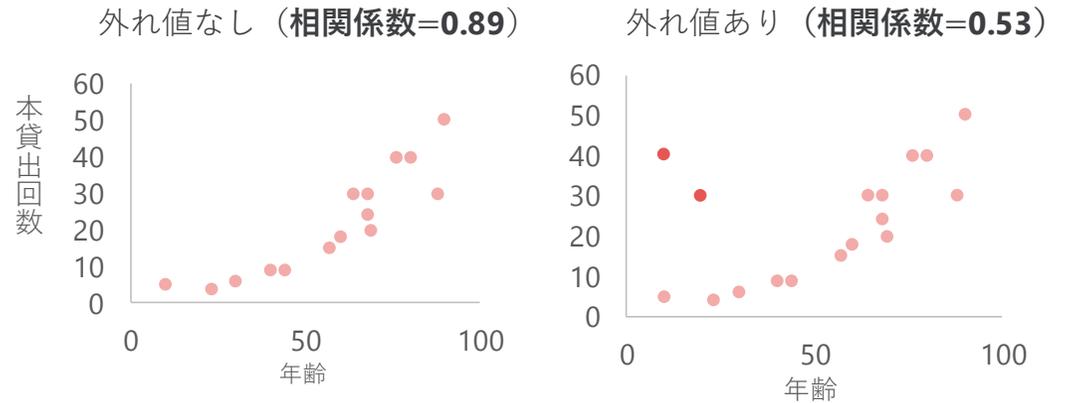
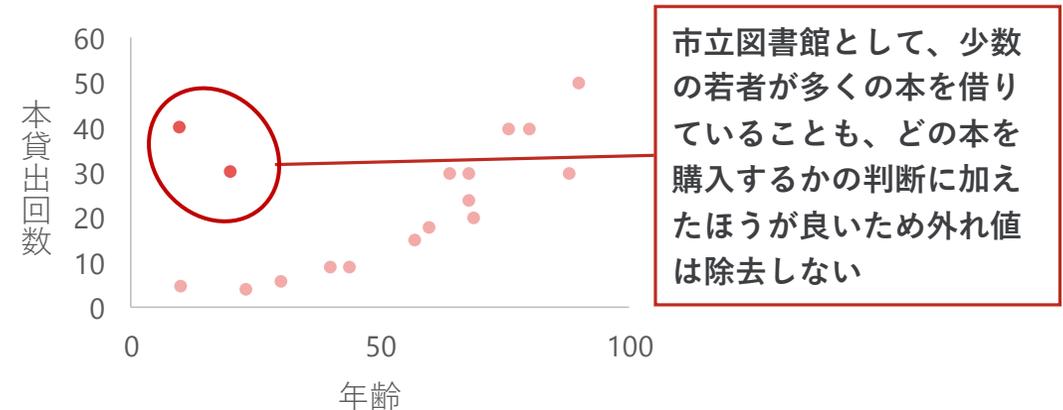


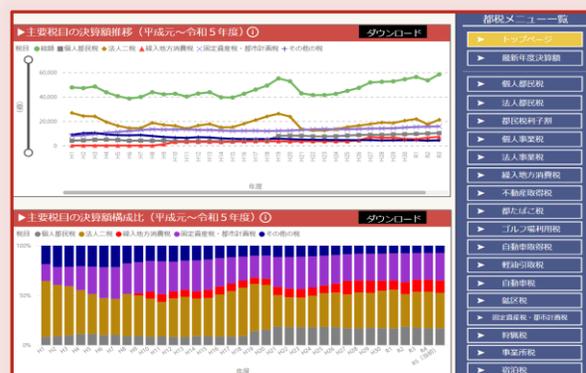
図2 外れ値を取り除かないほうが良い場合



## COLUMN データの可視化

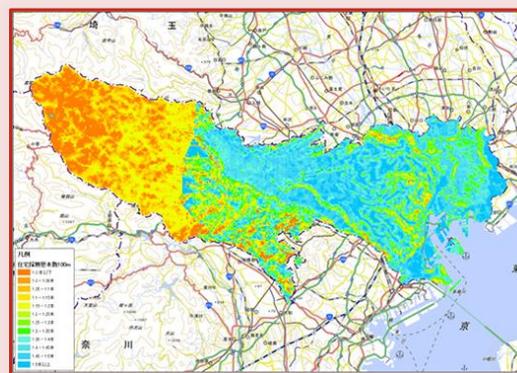
本ガイドラインでは仮説検証型の分析を主に説明していますが、グラフやマップでデータを視覚的に表現する方法（可視化）もあります。下記事例のように、提案や説明の説得力が増す、問題が分かり易くなる、情報共有や認識合わせが円滑になるなど、可視化は業務効率に大きな効果をもたらします。ぜひ、積極的に取り組みましょう。

※どのような観点から可視化すれば良いかについては、前述の「データを分析する観点」を参考にしてください。



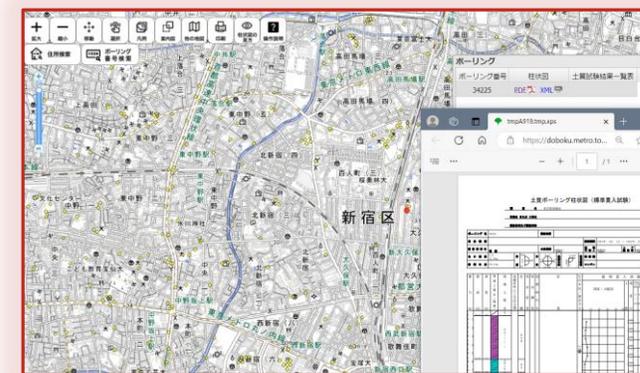
都税収入見える化ダッシュボード（主税局）

Power BIを用いて数値データを折れ線グラフや積み上げ棒グラフで表すことで、各税目の推移や決算額構成比などのデータの傾向が一目でわかるようになりました。



東京地中熱ポテンシャルマップ（産業労働局）

地中の熱利用のしやすさの違いを地図上に色別表示することで、都内における地中熱の採熱可能量（ポテンシャル）の目安が視覚的に把握できるようになりました。



東京の地盤（GIS版）（建設局）

地盤調査で得られた地質柱状図<sup>※1</sup>について、調査場所の位置情報を利用して地図上に配置することで、欲しいデータを検索・閲覧しやすくなりました。

※1：地質柱状図は、特定地域における地層の構造情報（岩質や厚さなど）を柱状に表現した図を指します

## 2 分析結果を考察する

### 1. Conclusionフェーズとは

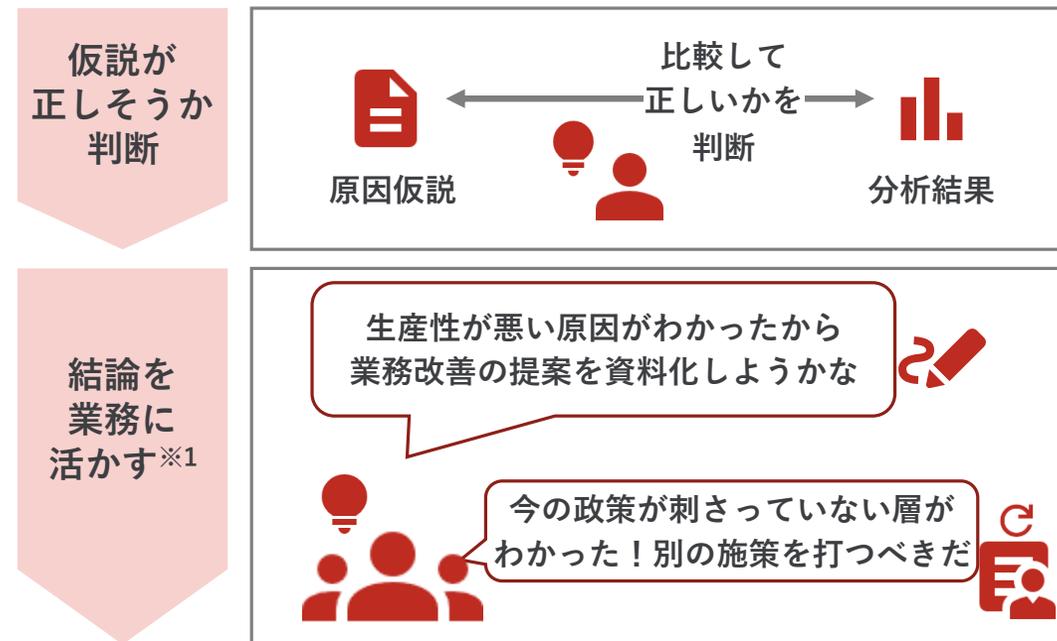
Analysisフェーズを通して有効な示唆を得られたとしても、示唆を得た時点で終わってしまっては事業として成果に結びつきません。

Conclusionフェーズでは、当初立てた原因仮説が正しかったかをAnalysisフェーズの結果から判断します。そのうえで、原因仮説が正しかった場合は分析の目的に立ち返り、解決策の提案や報告資料の作成など、結果を実際の業務に活かしていきましょう。

ワークシート更新

Conclusionエリア

#### Conclusionフェーズのステップ



Planフェーズでは想定内容を記載しましたが、Conclusionフェーズではこの記述を実際の結論に書き替えます。



#### 結論

まず「原因仮説」と「分析結果(見解)」を照合し、仮説を検証できたかを記入します。



#### 施策の方向性

原因を踏まえて次に提案する施策の方向や、検証する仮説を記入します。

※1：仮説が正しくなかった場合は別の仮説を検証したり課題を考え直しましょう

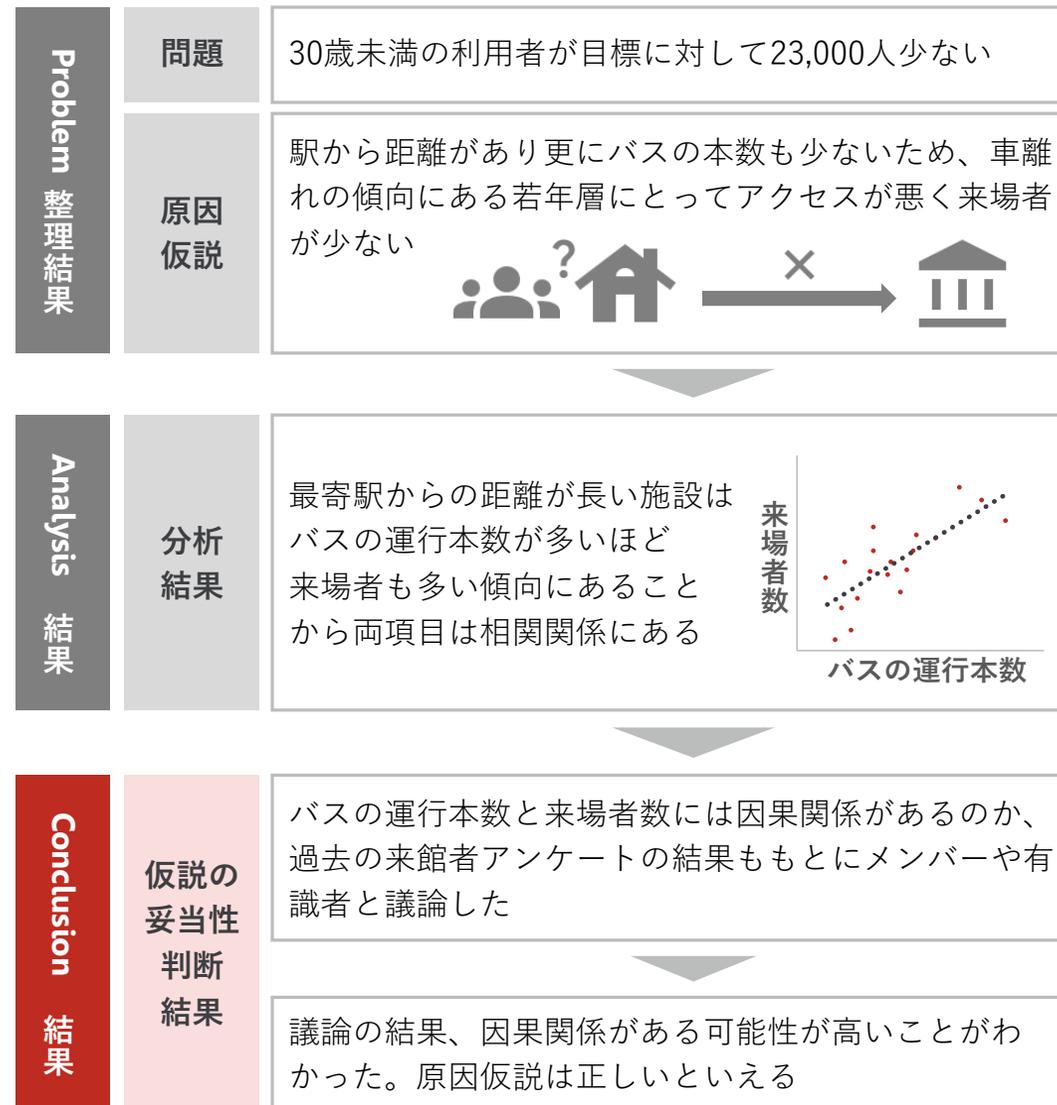
## 2 分析結果を考察する

### 2. 仮説が正しいか判断する

前述のとおり Conclusion フェーズでは、Problem フェーズで整理した原因仮説と Analysis フェーズの分析結果（見解）を比較することで、原因仮説が正しかったかどうか（妥当性）を判断します。

その際は、周囲の人の意見を聞きながら、客観的に考えることが重要です。例えば、職場のメンバーと論理に矛盾がないか確認したり、分析結果から他の示唆が得られないか議論する、などです。

なぜなら、これまで分析を計画し実行してきた本人だからこそ、気づかぬうちに自身の判断にバイアスがかかっている可能性があるためです。よくかかるバイアスとして、自分にとって都合の良い情報ばかりが目に入る確認バイアスが挙げられます。



### 3. 分析結果を業務に活かす

分析した結果をもって原因仮説が正しいと判断した場合は、当初の背景や問題に立ち返って分析結果を業務に活かしましょう。政策を提案する場合や資料を作成する際は、分析結果を用いて必ず根拠づけすることが重要です。数値などの定量データで根拠を示すことで説得力が増し、説明時間が短縮したり提案が通りやすくなるなどのメリットがあります。なお、このときも第三者に意見を聞いたり議論を行うようにしましょう。

一方で仮説が妥当でなかったとしても、「この原因は正しくないため、原因は他のところにあるようだ」という情報を得ることができます。分析結果から得られた新しい情報をより良い施策に繋げましょう。

#### 分析結果を業務に活かすイメージ



昨年の研修が職員のITリテラシー向上の大きく貢献したことがわかった。次年度の取組みの提案に アンケート結果を入れたら説得力が増す!

レク資料の現状分析ページは上司や有識者の方々も特に気にしていたな。 分析結果を使って根拠を説明すれば、短時間で理解しやすい!

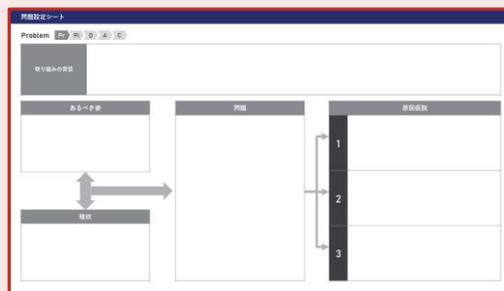


これまでは前年度を引き継ぐ形で政策を続けていたけど、別のやり方が効果的とわかった。 分析結果を使えば、それが一目でわかる!

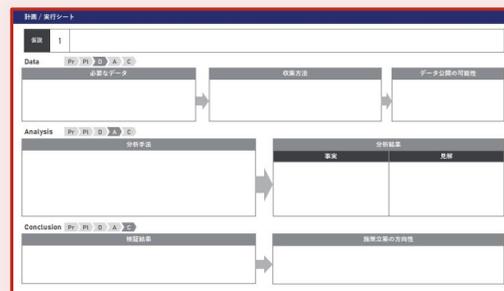
## 1. 分析のステップと結果の解釈

ここからは、企画編「ワークシートを活用して企画する」の記載例（下図）に対して実際に分析を実施し、結果を解釈するまでの流れを説明します。右図は、ワークシートの内容を簡略的に記載したものです。詳細な内容に関してはP.26,27の記入例をご覧ください。

問題設定シート



計画/実行シート



問題	30歳未満の利用者が目標に対して23,000人少ない
原因仮説	駅から距離があり更にバスの本数も少ないため、車離れの傾向にある若年層にとってアクセスが悪く来場者が少ない

本公共施設と類似している施設※1に対してデータ収集・分析を実施

必要なデータ※2	<ul style="list-style-type: none"> <li>① 最寄り駅から施設までの距離</li> <li>② 最寄り駅から施設までの1日のバス運行本数</li> <li>③ 30歳未満の1日あたりの平均来場者数</li> </ul>
分析方法	最寄り駅からの距離が長い施設※3を対象に、1日のバス運行本数と来場者数を散布図でグラフ化し、バスの運行本数が多ければ来場者数も多い傾向にあることを確認する。

※1: 施設の種類や大きさ、周辺人口が近い施設のこと。  
 ※2: 本例では、説明をしやすくするため仮想的に作成したデータを使用する。  
 ※3: 最寄り駅からの距離が1.6km（徒歩20分）以上を「最寄り駅から距離が長い施設」と仮置きする。

## 3 利活用の実践例

## 1. 分析のステップと結果の解釈

## 1 使用ツールの選択 (Power BI)

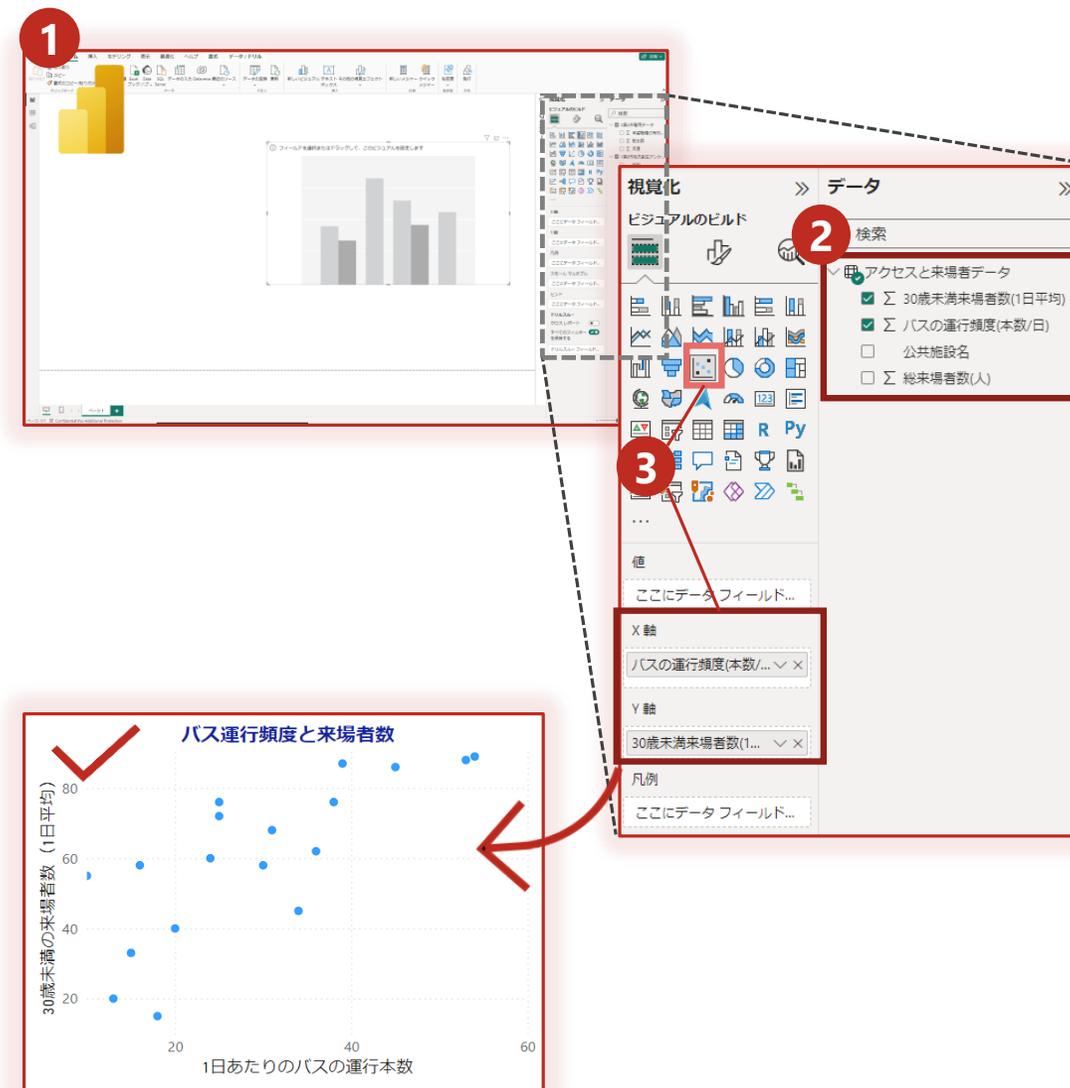
- 本実践例では分析のツールとして、**Power BI**を使用します。

## 2 データの選択

- 今回使用する「アクセスと来場者データ」を選択します。

## 3 データの表現方法の選択

- バスの1日あたりの運行数、30歳未満の来場者数の関連性を探りたいため、**散布図**を選択します。
- グラフでは「バスの運行頻度」と「30歳未満の1日あたりの平均来場者数」の分布を描きます。  
X軸→「バスの運行頻度 (本数/日)」を配置  
Y軸→「30歳未満来場者数 (1日平均)」を配置



# 1. 分析のステップと結果の解釈

## 4 分析の結果（事実）

バス運行本数と30歳未満の来場者数には以下の関係が見られました。

- バスの本数が20本以下→来場者数は70人以下
- バスの本数が21本~40本→来場者数は40人~90人 …etc

## 5 分析の結果（解釈）

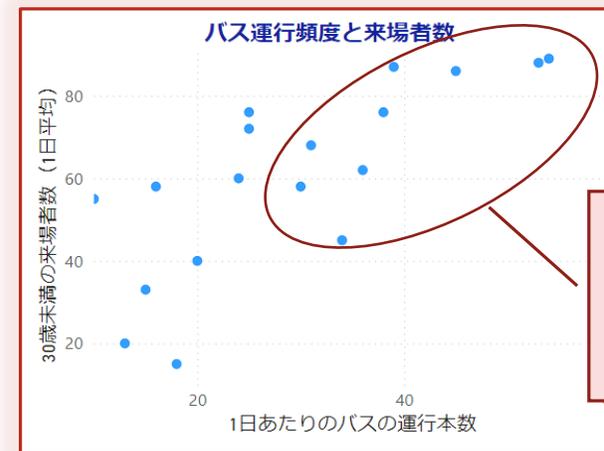
- 移動距離が長い場合、バスの運行本数が多ければ来場者数も多い傾向にあることから両項目は**相関関係にある**ことがわかります。

## 6 仮説が正しいかの判断

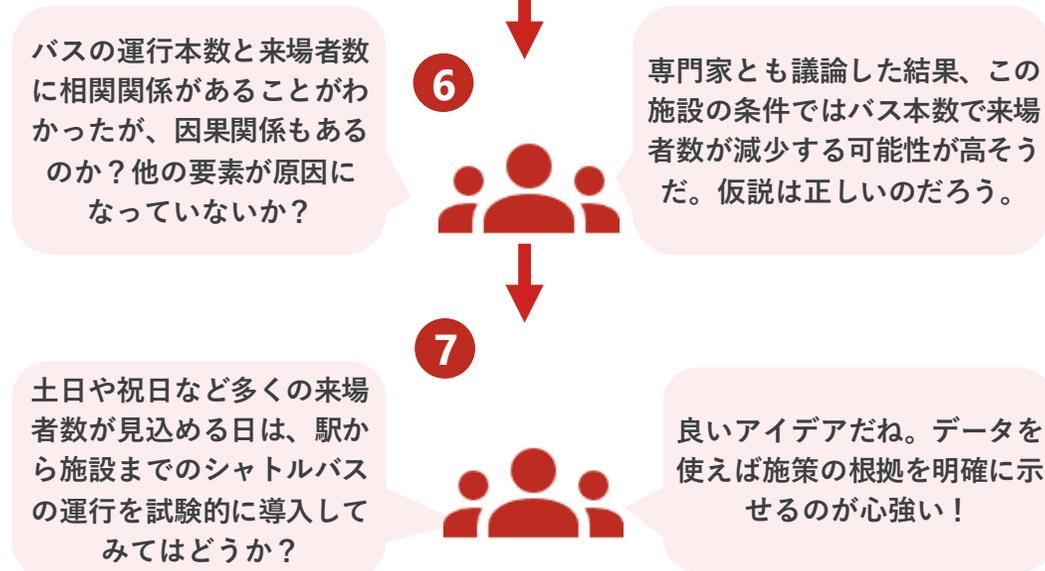
- 分析結果から得られた結果は合っているのか、**自問自答したり、メンバーや上司と議論した**うえで仮説の正しさを判断しましょう

## 7 分析結果を業務に活用

- 仮説が正しいことがわかったら、どのような施策を打つべきか、**チーム内で話し合**いましょう



バスを多めに運行している施設は来場者数が多い傾向にある



はじめに

企画

収集

利活用

チェックリスト

# チェックリスト

チェックリスト編では、収集編で説明した利用許諾取得とデータの品質管理について、業務で実践するためのチェックリストとその使用方法を紹介します。

本ガイドラインで紹介するチェックリストは、

「①利用許諾チェックリスト」と「②データ品質チェックリスト」の2点です。

各々のチェックリストが利用されるシーンやそれぞれのチェック観点の違いを把握し、業務で必要になった場合に本チェックリストを使えるようになりましょう。

- |   |              |       |      |
|---|--------------|-------|------|
| ① | 利用許諾チェックリスト  | ————— | P.60 |
| ② | データ品質チェックリスト | ————— | P.61 |

## 1 利用許諾チェックリスト

## 1. 利用許諾チェックリストの概要

利用許諾を取得する際は、経済産業省が発行している

「データ利活用チェックシート」※1を活用します。

本チェックリストは契約やセキュリティなどの観点から確認すべき重要事項を①データ提供、②データ取得・保有、③データ使用、④プラットフォームの4つのシーンに分けて整理しています。本ガイドラインは主として①、②、③での活用を想定しています。

チェックリストから自分の状況に合致した項目を参照し、データの利活用やデータ公開などが可能か確認するようにしましょう。



ワンポイント  
利用許諾取得の考え方については  
収集編 P.32~34を参照ください

## 利用許諾チェックリストの利用シーン

	利用シーン	チェック項目（例）
データ提供	データを組織外に提供する 	<ul style="list-style-type: none"> <li>✓ 契約で目的外使用の禁止などを規定する</li> <li>✓ 個人情報の第三者提供の可否を確認する</li> </ul>
データ取得/保有	データを収集して管理する 	<ul style="list-style-type: none"> <li>✓ データ提供者に対して、データに個人情報が含まれていないことの保証を要求する</li> </ul>
データ使用	データを使用する 	<ul style="list-style-type: none"> <li>✓ 著作権や著作人格権を侵害しないよう留意のうえデータを取り扱う</li> <li>✓ 契約における利用範囲を確認する</li> </ul>
プラットフォーム	データを共有する 	<ul style="list-style-type: none"> <li>✓ 設計に際し、中立性・信頼性の確保を意識する</li> </ul>

※1：「データ利活用チェックシート」は、経済産業省の「データ利活用のポイント集」（巻末付録：リンク一覧⑥）を参照してください。

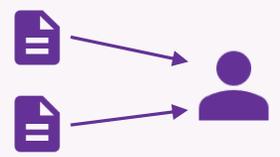
# 1. データ品質チェックリストの概要

データ品質管理を実践する際は、都で作成した「データ品質チェックリスト」※1を活用します。

「データ品質チェックリスト」は、データ品質を高めるためのチェック項目を「企画」「要件定義」「設計・開発」の工程ごとに用意しています。これまで説明してきたようなPPDACに沿ってデータを収集する場面はもちろん、システムの要件定義や設計・開発といった場面でも活用できます。

次節以降では、工程別に具体的なチェックリストの利用シーンや使用方法を説明します。なお、チェックリストは付録を参照ください。

データ品質チェックリストの利用シーン

	利用シーン	チェック項目の概要
<b>企画</b> P.62	PPDACに沿ってデータ利活用を進める場合  データを収集する	収集したデータに対し、実施すべきアクション項目
<b>要件定義</b> P.65	システム開発など利活用されるデータを生成する場合  仕様書を作成する	データ品質に関連する要件として考慮すべき観点
<b>設計/開発</b> P.66	システムを設計する 	データに関連する設計内容に対し実施すべきアクション項目

※1：デジタル庁の「データ品質管理ガイドブック」（巻末付録：リンク一覧④）を基に作成しました。

## ② データ品質チェックリスト

フェーズ

Pr

Pl

D

A

C

Data

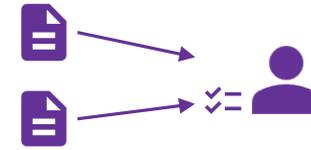
## 2. 使用方法（企画工程）

企画工程向けのチェックリストは、PPDACに沿ってデータ利活用を行う場合、特にDataフェーズでデータを収集する際に使用します。データ収集の場面としては、以下2つがあります。

1つは、データ利活用に際し職員が調査票やアンケートなどでデータを収集する場面です。職員はGoogle FormsやMicrosoft Formsなどのツールで調査票やアンケートなどを作成したあと、チェックリストを使ってデータ品質のチェック項目を満たしているか確認します。

もう1つは、職員がデータ収集を事業者に委託する場面です。職員は仕様書を作成する際に、チェックリストの項目をデータ品質の要件として掲載します。

### 利用シーン



データを収集する際に実施すべきアクション項目を確認する



データ収集を事業者に委託する際の仕様書にデータ品質要件を盛り込む

### 企画工程のチェックリスト（抜粋）

#### チェック項目（例）

#### 正確性

✓ アンケートや調査票を作成する際に、書式を指定できるところについて、入力欄の書式を指定しているか

#### 完全性

✓ アンケートや調査票を作成する際に、利用目的と照らして必要なデータ項目が網羅されているか

#### 一貫性

✓ 同一ファイル内でデータの矛盾はないか

#### 可用性

✓ 必要なときにいつでもデータにアクセスできるようになっているか

#### 最新性

✓ 公開データの更新サイクルが元データの更新サイクルに対して適切に設定されているか



## Excelでデータを作るときの心掛け (1/2) ※1

一般的に、データを収集するときにはExcelを使うことが多いでしょう。その際は、データを利活用しやすいように機械判読性が高い形式に保つことが重要です。以下の項目を守れているか確認しましょう。

### 1つのセルには1データ

1セルにデータが複数あると、後で利用しにくいから控えましょう。

×	○																											
<table border="1"> <thead> <tr> <th>項番</th> <th>避難場所</th> </tr> </thead> <tbody> <tr> <td>港区</td> <td>芝浦小学校 港南小学校 港南中学校</td> </tr> <tr> <td>新宿区</td> <td>天神小学校 四谷小学校 花園小学校</td> </tr> </tbody> </table>	項番	避難場所	港区	芝浦小学校 港南小学校 港南中学校	新宿区	天神小学校 四谷小学校 花園小学校	<table border="1"> <thead> <tr> <th>項番</th> <th>自治体</th> <th>場所</th> </tr> </thead> <tbody> <tr> <td>XXX</td> <td>港区</td> <td>芝浦小学校</td> </tr> <tr> <td>XXX</td> <td>港区</td> <td>港南小学校</td> </tr> <tr> <td>XXX</td> <td>港区</td> <td>港南中学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>天神小学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>四谷小学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>花園小学校</td> </tr> </tbody> </table>	項番	自治体	場所	XXX	港区	芝浦小学校	XXX	港区	港南小学校	XXX	港区	港南中学校	XXX	新宿区	天神小学校	XXX	新宿区	四谷小学校	XXX	新宿区	花園小学校
項番	避難場所																											
港区	芝浦小学校 港南小学校 港南中学校																											
新宿区	天神小学校 四谷小学校 花園小学校																											
項番	自治体	場所																										
XXX	港区	芝浦小学校																										
XXX	港区	港南小学校																										
XXX	港区	港南中学校																										
XXX	新宿区	天神小学校																										
XXX	新宿区	四谷小学校																										
XXX	新宿区	花園小学校																										

### 結合はやめよう

複数のセルを結合するのも、使いにくい原因になります。

×	○																																										
<table border="1"> <thead> <tr> <th>項番</th> <th>自治体</th> <th>場所</th> </tr> </thead> <tbody> <tr> <td>XXX</td> <td></td> <td>芝浦小学校</td> </tr> <tr> <td>XXX</td> <td>港区</td> <td>港南小学校</td> </tr> <tr> <td>XXX</td> <td></td> <td>港南中学校</td> </tr> <tr> <td>XXX</td> <td></td> <td>天神小学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>四谷小学校</td> </tr> <tr> <td>XXX</td> <td></td> <td>花園小学校</td> </tr> </tbody> </table>	項番	自治体	場所	XXX		芝浦小学校	XXX	港区	港南小学校	XXX		港南中学校	XXX		天神小学校	XXX	新宿区	四谷小学校	XXX		花園小学校	<table border="1"> <thead> <tr> <th>項番</th> <th>自治体</th> <th>場所</th> </tr> </thead> <tbody> <tr> <td>XXX</td> <td>港区</td> <td>芝浦小学校</td> </tr> <tr> <td>XXX</td> <td>港区</td> <td>港南小学校</td> </tr> <tr> <td>XXX</td> <td>港区</td> <td>港南中学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>天神小学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>四谷小学校</td> </tr> <tr> <td>XXX</td> <td>新宿区</td> <td>花園小学校</td> </tr> </tbody> </table>	項番	自治体	場所	XXX	港区	芝浦小学校	XXX	港区	港南小学校	XXX	港区	港南中学校	XXX	新宿区	天神小学校	XXX	新宿区	四谷小学校	XXX	新宿区	花園小学校
項番	自治体	場所																																									
XXX		芝浦小学校																																									
XXX	港区	港南小学校																																									
XXX		港南中学校																																									
XXX		天神小学校																																									
XXX	新宿区	四谷小学校																																									
XXX		花園小学校																																									
項番	自治体	場所																																									
XXX	港区	芝浦小学校																																									
XXX	港区	港南小学校																																									
XXX	港区	港南中学校																																									
XXX	新宿区	天神小学校																																									
XXX	新宿区	四谷小学校																																									
XXX	新宿区	花園小学校																																									

### 数値と文字列は分ける

数値として文字は混ぜるな危険。文字列として認識されます。

×	○																																										
<table border="1"> <thead> <tr> <th>施設名</th> <th>時間</th> <th>金額</th> </tr> </thead> <tbody> <tr> <td>XXX</td> <td>2時間</td> <td>2千円</td> </tr> <tr> <td>XXX</td> <td>2</td> <td>2,000</td> </tr> <tr> <td>XXX</td> <td>2</td> <td>500円</td> </tr> <tr> <td>XXX</td> <td>1:00</td> <td>300</td> </tr> <tr> <td>XXX</td> <td>1時間</td> <td>200</td> </tr> <tr> <td>XXX</td> <td>1</td> <td>壹万円</td> </tr> </tbody> </table>	施設名	時間	金額	XXX	2時間	2千円	XXX	2	2,000	XXX	2	500円	XXX	1:00	300	XXX	1時間	200	XXX	1	壹万円	<table border="1"> <thead> <tr> <th>施設名</th> <th>時間</th> <th>金額</th> </tr> </thead> <tbody> <tr> <td>XXX</td> <td>2</td> <td>2000</td> </tr> <tr> <td>XXX</td> <td>2</td> <td>2000</td> </tr> <tr> <td>XXX</td> <td>2</td> <td>500</td> </tr> <tr> <td>XXX</td> <td>1</td> <td>300</td> </tr> <tr> <td>XXX</td> <td>1</td> <td>200</td> </tr> <tr> <td>XXX</td> <td>1</td> <td>10000</td> </tr> </tbody> </table>	施設名	時間	金額	XXX	2	2000	XXX	2	2000	XXX	2	500	XXX	1	300	XXX	1	200	XXX	1	10000
施設名	時間	金額																																									
XXX	2時間	2千円																																									
XXX	2	2,000																																									
XXX	2	500円																																									
XXX	1:00	300																																									
XXX	1時間	200																																									
XXX	1	壹万円																																									
施設名	時間	金額																																									
XXX	2	2000																																									
XXX	2	2000																																									
XXX	2	500																																									
XXX	1	300																																									
XXX	1	200																																									
XXX	1	10000																																									

### 不要な空白/改行はNG

スペースや改行は、正しくデータが読み取れない原因になります。

×	○																																				
<table border="1"> <thead> <tr> <th>施設名</th> <th>時間</th> <th>金額</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>2</td> <td>2000</td> </tr> <tr> <td>B</td> <td>2</td> <td>2000</td> </tr> <tr> <td>C</td> <td>1</td> <td>500</td> </tr> <tr> <td>会 場</td> <td>1</td> <td>500</td> </tr> <tr> <td>C</td> <td>1</td> <td>300</td> </tr> <tr> <td>音 響</td> <td>1</td> <td>300</td> </tr> </tbody> </table>	施設名	時間	金額	A	2	2000	B	2	2000	C	1	500	会 場	1	500	C	1	300	音 響	1	300	<table border="1"> <thead> <tr> <th>施設名</th> <th>時間</th> <th>金額</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>2</td> <td>2000</td> </tr> <tr> <td>B</td> <td>2</td> <td>2000</td> </tr> <tr> <td>C会場</td> <td>1</td> <td>500</td> </tr> <tr> <td>C音響</td> <td>1</td> <td>300</td> </tr> </tbody> </table>	施設名	時間	金額	A	2	2000	B	2	2000	C会場	1	500	C音響	1	300
施設名	時間	金額																																			
A	2	2000																																			
B	2	2000																																			
C	1	500																																			
会 場	1	500																																			
C	1	300																																			
音 響	1	300																																			
施設名	時間	金額																																			
A	2	2000																																			
B	2	2000																																			
C会場	1	500																																			
C音響	1	300																																			

スペースや改行で体裁を整えている

スペースや改行を解除した状態

出典：統計表における機械判読可能なデータ 作成に関する表記方法（総務省）

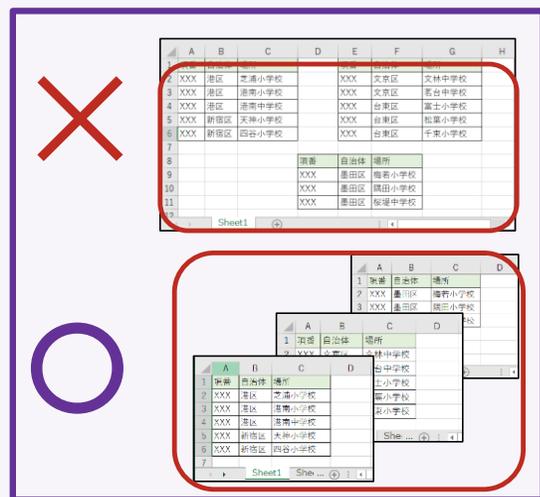
出典：Tips for creating machine-friendly data（内田康雄）

※1：東京都の「データ整備の必要性とマニュアルの使い方」にも同内容が記載されています

Excelでデータを作るときにの心掛け (2/2) ※1

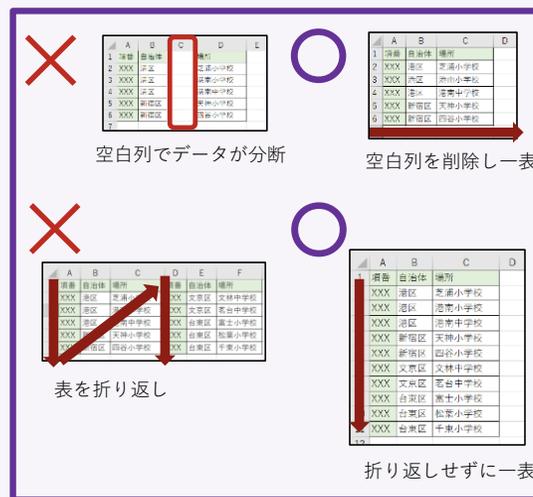
1シート1テーブル

扱いやすいように1シートに表は1つ。複数あるときは分割しましょう。



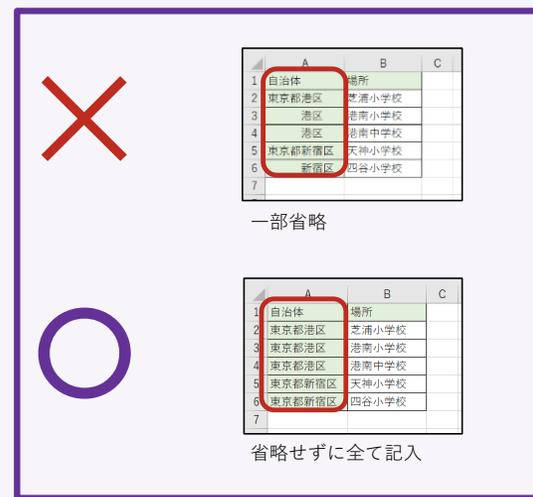
データを分断しない

不必要な表の分離は機械判読を阻害します。



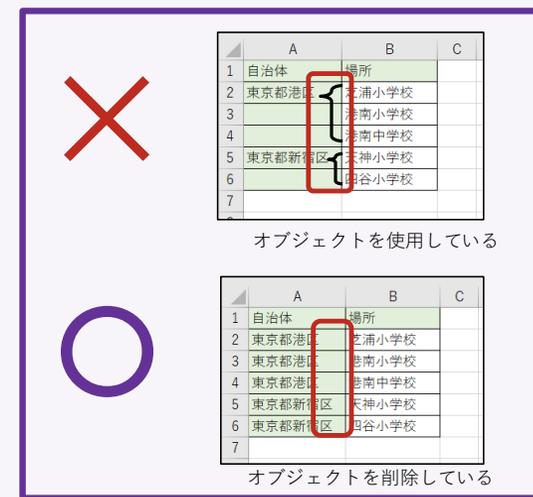
項目名はそのままに

項目名が省略されると機械が自動で判読できなくなります。



オブジェクトは使わない

オブジェクトは扱いにくいいため、セル入力を基本としましょう。



Excelを活用してデータを作成した場合は、CSVファイルでの保管も検討しましょう。CSVファイルは、各項目がカンマで区切られたテキストデータで、データのサイズが小さく機械判読が容易という特徴があります。

出典：統計表における機械判読可能なデータ作成に関する表記方法（総務省）  
 出典：Tips for creating machine-friendly data（内田康雄）  
 ※1：東京都の「データ整備の必要性和マニュアルの使い方」にも同内容が記載されています

### 3. 使用方法（要件定義工程）

要件定義工程向けのチェックリストは、システムの設計・開発工程の仕様書を作成する際に使用します。

調達の背景や状況を踏まえて、チェックリストを参考にデータ品質要件を仕様書に掲載することで、特に都が重視しているデータ品質の考えを示すことができます。

仕様書を作成する際には、チェックリストにあるデータ品質の5つの観点を記載することが重要です。すべてチェックリスト上の例示の通り記載するのではなく、開発するシステムの機能要件や仕様を踏まえ、データ品質要件の記載を適宜調整しましょう。

#### 利用シーン

仕様書にデータ品質の5つの観点が盛り込まれているか確認してみよう



チェックリストを活用して仕様書にデータ品質要件を盛り込む

#### 要件定義工程のチェックリスト（抜粋）

##### 要件定義工程のチェック項目（例）

##### 正確性

✓ 利用者が入力するデータを間違えないよう、バリデーションチェックなどの仕組みを設定する

##### 完全性

✓ データの入力漏れを制御するなど、利用目的の達成に必要なデータが漏れなく入力されるようにする

##### 一貫性

✓ 入力データに矛盾がないよう制御するなど、システム内および東京都の既存システムとの連携においてデータの一貫性を保てるようにする

##### 可用性

✓ データを常に利用可能な状態にする

##### 最新性

✓ データの利用目的に適した更新頻度で組織内又は組織外に公開するデータを更新する

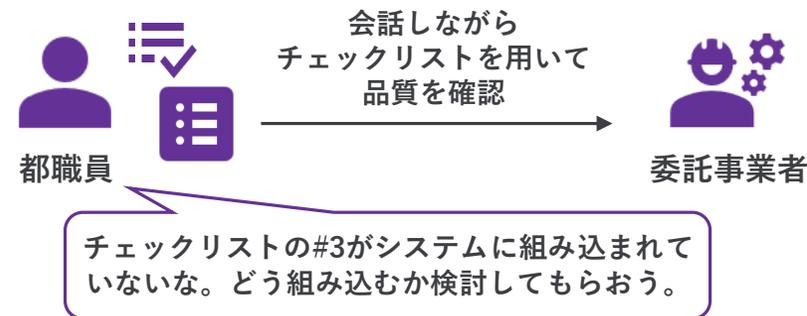
## 4. 使用方法（設計・開発工程）

設計・開発工程向けのチェックリストは、要件定義や仕様書で定めたデータ品質維持に必要な要件が適切に設計・開発に組み込んでいるか、確認する際に活用します。

設計・開発工程は委託事業者任せにするのではなく、チェックリストを用いて委託事業者と会話しながら確認することが重要です。高いデータ品質を維持できる仕組みになっているかを共に確認することで、システムのデータ品質を一定以上に担保できます。

このときチェック対象とする項目は、要件定義や仕様、システムの特性と合致するものを選ぶようにしましょう。

### 利用シーン



### 設計・開発工程のチェックリスト（抜粋）

#### 設計・開発工程のチェック項目（例）

正確性

✓ 書式を指定できるものについて、入力欄の書式を制御しているか

完全性

✓ 利用目的と照らしたときに必要なデータ項目が網羅されているか

一貫性

✓ システム内で一意に主キーを設定しているか

可用性

✓ 必要なときにいつでもデータにアクセスできるようになっているか

最新性

✓ 公開データの更新サイクルが元データの更新サイクルに対して適切に設定されているか

## 巻末付録：リンク一覧

- ① 総務省「Data StaRt」セミナー編（1）～データ利活用の進め方（<https://www.stat.go.jp/dstart/point/seminar1/>）2024年3月13日閲覧
- ② デジタル庁「入力フォーム構築時の指針となるテンプレートを追加しました」（<https://www.digital.go.jp/policies/servicedesign/designsystem/20230531-1>）2024年3月13日閲覧
- ③ 高度情報通信ネットワーク社会推進戦略本部（2017）「オープンデータ基本指針」（[https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/f7fde41d-ffca-4b2a-9b25-94b8a701a037/20210615\\_resources\\_data\\_guideline\\_01.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/f7fde41d-ffca-4b2a-9b25-94b8a701a037/20210615_resources_data_guideline_01.pdf)）2024年3月13日閲覧
- ④ デジタル庁（2022）「データ品質管理ガイドブック」（[https://github.com/JDA-DM/GIF/blob/main/460\\_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF/md/468-1\\_guidebook\\_dataquality.md](https://github.com/JDA-DM/GIF/blob/main/460_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF/md/468-1_guidebook_dataquality.md)）2024年3月13日閲覧
- ⑤ 情報通信技術（IT）総合戦略室（2021）「ベース・レジストリの指定について」（<https://cio.go.jp/sites/default/files/uploads/documents/SpecifyingBaseRegistry.pdf>）2024年3月13日閲覧
- ⑥ 経済産業省（2020）「データ利活用のポイント集」（<https://www.meti.go.jp/policy/economy/chizai/chiteki/pdf/datapoint.pdf>）2024年3月13日閲覧

## 巻末付録：本文で紹介している東京都のサービス・ツール

ページ	サービス・ツール名称	URL
P6	東京都デジタルサービスの開発・運用に係る行動指針	<a href="https://www.digitalservice.metro.tokyo.lg.jp/documents/d/digitalservice/02_koudoushishin">https://www.digitalservice.metro.tokyo.lg.jp/documents/d/digitalservice/02_koudoushishin</a>
P19	東京都デジタルツイン実現プロジェクト	<a href="https://info.tokyo-digitaltwin.metro.tokyo.lg.jp/">https://info.tokyo-digitaltwin.metro.tokyo.lg.jp/</a>
P31	東京オープンデータカタログサイト	<a href="https://portal.data.metro.tokyo.lg.jp/">https://portal.data.metro.tokyo.lg.jp/</a>
	東京データプラットフォーム	<a href="https://www.tdpf.metro.tokyo.lg.jp/">https://www.tdpf.metro.tokyo.lg.jp/</a>
P52	東京の地盤（GIS版）	<a href="https://doboku.metro.tokyo.lg.jp/start/03-jyouhou/geo-web/geo-webmap.aspx">https://doboku.metro.tokyo.lg.jp/start/03-jyouhou/geo-web/geo-webmap.aspx</a>
	東京地中熱ポテンシャルマップ	<a href="https://www.tokyogeoheatmap.metro.tokyo.lg.jp/">https://www.tokyogeoheatmap.metro.tokyo.lg.jp/</a>
	都税収入見える化ダッシュボード	<a href="https://app.powerbi.com/view?r=eyJrljoiZjUzNmVjMTAtMjA0Ni00ODI0LW11ODctMDI2MWEzNDlyMThkliwidCI6ImQwMzAyZmNjLTNIODEtNDljMy04MjM1LWQzMjFhMzY4NGNmYyJ9&amp;pageName=ReportSection862d2d37bcc6ca2a0120">https://app.powerbi.com/view?r=eyJrljoiZjUzNmVjMTAtMjA0Ni00ODI0LW11ODctMDI2MWEzNDlyMThkliwidCI6ImQwMzAyZmNjLTNIODEtNDljMy04MjM1LWQzMjFhMzY4NGNmYyJ9&amp;pageName=ReportSection862d2d37bcc6ca2a0120</a>
P64,65	データ整備の必要性和マニュアルの使い方	<a href="https://www.digitalservice.metro.tokyo.lg.jp/documents/d/digitalservice/houkokukai_01">https://www.digitalservice.metro.tokyo.lg.jp/documents/d/digitalservice/houkokukai_01</a>

## その他参考文献

- 内田康雄「Tips for creating machine-friendly data」 (<https://www.slideshare.net/uchiruda/tips-for-creating-machine-friendly-data-248311530>) 2024年3月19日閲覧
- 江崎貴裕 (2020) 「分析者のためのデータ解析学 入門 データの本質をとらえる技術」 ソシム
- 気象庁「気温データ年ごとの値」 ([https://www.data.jma.go.jp/obd/stats/etrn/view/annually\\_s.php?prec\\_no=44&block\\_no=47662](https://www.data.jma.go.jp/obd/stats/etrn/view/annually_s.php?prec_no=44&block_no=47662)) 2024年3月19日閲覧
- 総務省「EBPMブートキャンプ<データ分析ワークブック>」 ([https://www.stat.go.jp/dstart/research/case\\_ebpm/pdf/workbook.pdf](https://www.stat.go.jp/dstart/research/case_ebpm/pdf/workbook.pdf)) 2024年3月19日閲覧
- 総務省「統計表における機械判読可能なデータ作成に関する表記方法」 ([https://www.soumu.go.jp/main\\_content/000723626.pdf](https://www.soumu.go.jp/main_content/000723626.pdf)) 2024年3月19日閲覧
- 総務省「就業構造基本調査」 ([https://www.e-stat.go.jp/stat-search/database?page=1&toukei=00200532&result\\_page=1](https://www.e-stat.go.jp/stat-search/database?page=1&toukei=00200532&result_page=1)) 2024年3月11日閲覧
- 総務省「地方公共団体における行政改革の取組」 ([https://www.soumu.go.jp/main\\_content/000880862.pdf](https://www.soumu.go.jp/main_content/000880862.pdf)) 2024年3月11日閲覧
- 総務省「地方公共団体におけるデータ利活用ガイドブック」 ([https://www.soumu.go.jp/main\\_content/000620312.pdf](https://www.soumu.go.jp/main_content/000620312.pdf)) 2024年3月11日閲覧
- 総務省「行政におけるデータ利活用の推進」 ([https://www.kantei.go.jp/jp/singi/toukeikaikaku/toukeigijutsu\\_data\\_source/dai4/siryou1.pdf](https://www.kantei.go.jp/jp/singi/toukeikaikaku/toukeigijutsu_data_source/dai4/siryou1.pdf)) 2024年3月11日閲覧
- 総務省「e-Stat 政府統計の総合窓口」 (<https://www.e-stat.go.jp/>) 2024年3月15日閲覧
- デジタル庁 (2023) 「デジタル・ガバメント推進標準ガイドライン」 ([https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/8a3b6203/20230331\\_resources\\_standard\\_guidelines\\_guideline\\_01.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/8a3b6203/20230331_resources_standard_guidelines_guideline_01.pdf)) 2024年3月15日閲覧
- デジタル庁「e-Govポータル」データポータル (<https://data.e-gov.go.jp/info/ja>) 2024年3月15日閲覧
- 内閣府、経済産業省「RESAS 地域経済分析システム」 (<https://resas.go.jp/#/13/13101>) 2024年3月15日閲覧
- 富士通総研 (2020) 「地方公共団体のデータ利活用に関する調査」 ([https://www.chisou.go.jp/sousei/resas/pdf/r2\\_03\\_houkokusho\\_fujitsu.pdf](https://www.chisou.go.jp/sousei/resas/pdf/r2_03_houkokusho_fujitsu.pdf)) 2024年3月15日閲覧
- マクロミル、渋谷智之 (2022) 「データ利活用の教科書」 翔泳社