

データ利活用 ガイドライン

Version 2.0.0

2025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及び目的 P.4
- 2 ガイドラインの構成 P.8

第一章：データを可視化する

- 1 データの可視化の定義と重要性 P.14
- 2 データの可視化の進め方 P.16

第二章：データを分析する

- 1 データ分析の概要と進め方 P.20
- 2 問題を特定し仮説を構築する (Problem) P.23

- 3 仮説検証のための計画を立てる (Plan) P.27
- 4 データを収集する (Data) P.31
- 5 データを分析する (Analysis) P.40
- 6 分析結果を考察する (Conclusion) P.51
- 7 データ分析の実践例 P.55

第三章：チェックリスト

- 1 利用許諾チェックリスト P.61
- 2 データ品質チェックリスト P.62

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理解しておきたいことを紹介します。

- 1 ガイドラインの背景及び目的 P.4
- 2 ガイドラインの構成 P.8

1 ガイドラインの背景及び目的

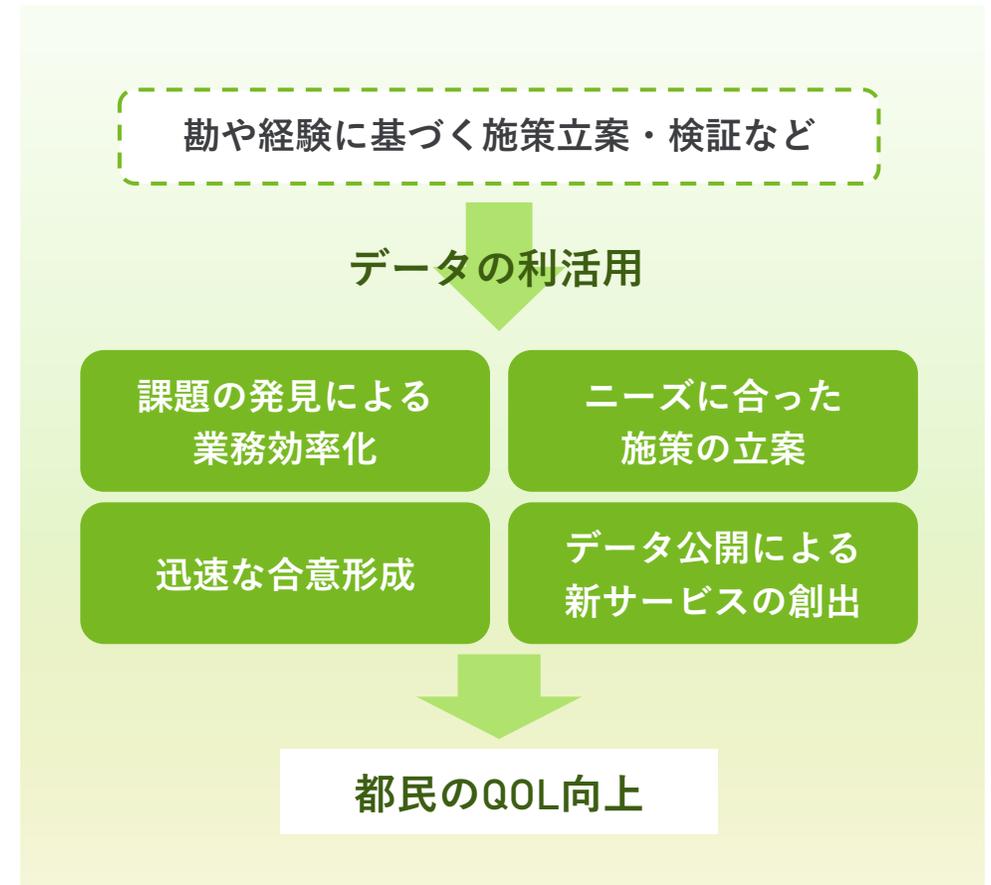
1. 本ガイドライン策定の背景

現在の日本では、人々の価値観の多様化や行政課題の複雑高度化に加えて少子高齢化など、行政を取り巻く社会経済状況が大きく変化しています。都庁においても、今後は限られた人的資源の中で業務を行うことが求められる可能性があります。

このような状況下で都民ニーズを的確に把握するとともに、効果的に業務を行う手段として、データの利活用が重要になります。日々の業務においてデータに基づいて確認し判断することで、効率的な課題発見や迅速な合意形成に繋がるほか、ニーズに合った施策の立案を行うことが可能になります。

職員ひとりひとりがデータの利活用を意識することで、都政のQOS向上、ひいては都民のQOL向上に繋がります。

データ利活用で期待される業務の変化



1 ガイドラインの背景及び目的

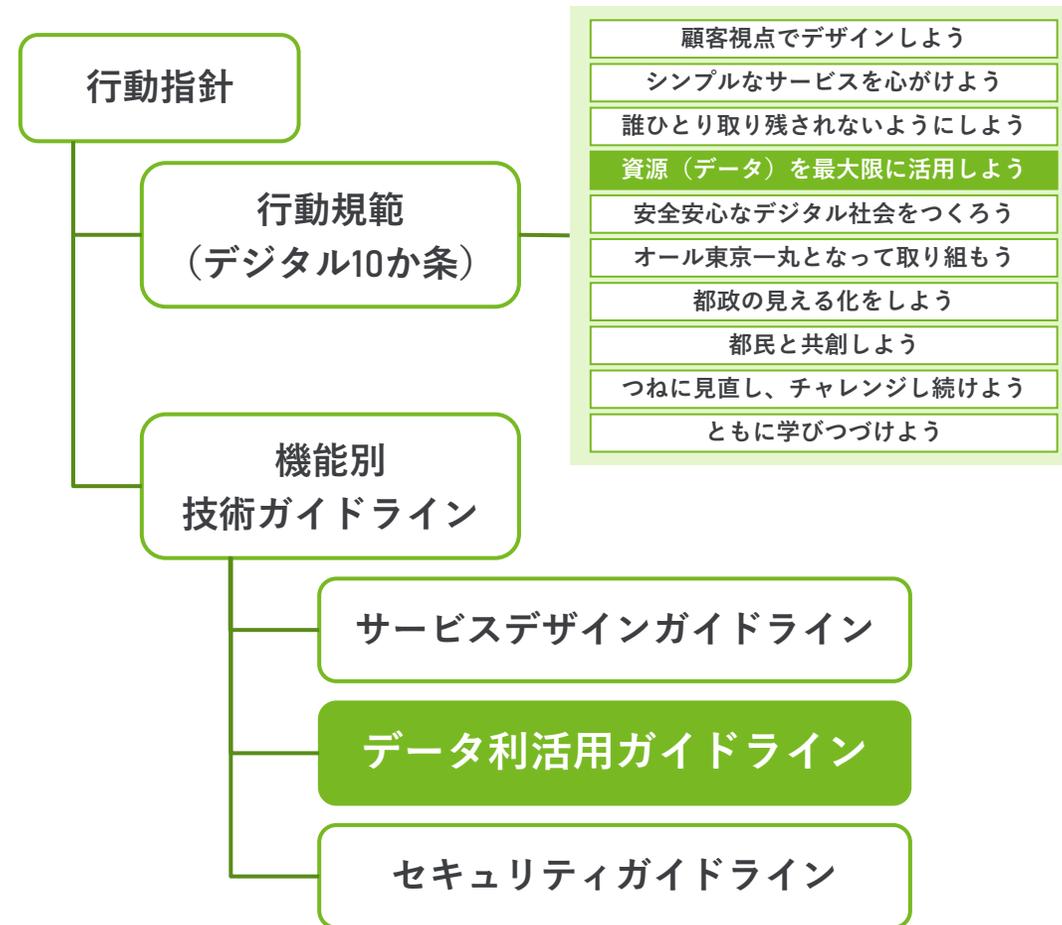
2. 本ガイドラインの位置づけ

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、東京都では2022年3月に「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。行動規範（デジタル10か条）には、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、10個の共通価値観が示されています。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



1 ガイドラインの背景及び目的

3. 本ガイドラインの目的

データ利活用が重要と理解しつつも、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、職員の皆さんに「自分の業務でもデータを利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、実践の機会が増え、さらに理解が高まっていくでしょう。この好循環が行政サービスの品質を向上させることに繋がります。

本ガイドラインでは、データ利活用を進める際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。

データ利活用の実践による好循環



1 ガイドラインの背景及び目的

4. 本ガイドラインの利用シーン

本ガイドラインは職員が日々の業務のなかでデータ活用を企画し実行する場面での利用を想定しています。

例えば、

- 事業や施策の検討における課題特定
- システムやツールの設計におけるデータ要件検討
- 施策実施時のKPIや成果指標の測定

など、業務における様々な場面が想定されます。

本ガイドラインは、所属部署や担当業務の種別に関係なく全職員を対象とし、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。

また末尾では、利用許諾の取得やデータ品質の管理の際に活用できるチェックリストを掲載しています。是非こちらも参考にしてみてください。

利用シーンの例



データを確認して
現状の課題を
特定しよう

設計中のシステムにおいて
どのようにデータ要件を
定義すれば使いやすくなるかな



ワンポイント

業務におけるデータ利活用には様々な形があります。例えば、以下の3つの例は取組の規模やデータ量も様々ですが、いずれも「データ利活用」に該当します。

例

- データをグラフ化し、根拠として報告資料に掲載する
- 集計した過去のデータからパターンを見つけ、施策検討に利用する
- 他部署のデータも確認できるよう、データ共有のルールを整備する

「データ利活用」は、必ずしも技術的や規模的に難易度が高いというわけではありません。まずはご自身の身近な業務の中でデータに触れてみるるところから始めましょう。

2 ガイドラインの構成

1. 本ガイドラインで扱うデータ利活用の類型

データ利活用には様々な類型が存在しますが、本ガイドラインでは主にデータの可視化と分析について説明しています。※1

可視化とは、データをグラフやチャートなどの視覚的な形式で表現し、情報を直感的に理解しやすくするプロセスです。様々なデータ利活用の類型の中で最も身近で分かりやすく、本ガイドラインでもデータ利活用のファーストステップと位置付けています。

分析とは、データを整理して起きている事象を把握し、その事象の原因や傾向などを明らかにして課題解決に繋げるためのプロセスです。業務において様々な意思決定や施策検討を行う際に役立つものであり、本ガイドラインでも重要な類型として位置付けています。

※1：可視化・分析以外にも、より発展的かつ専門的な知見・技術が求められる、検知や予測、最適化など様々なデータ利活用の類型があります。

※2：本ガイドラインで推奨しているアプローチは検証型と呼ばれるものですが、検証型以外にも発見型と呼ばれるアプローチも存在します。発見型は仮説を立てずに分析を行うアプローチであり、これは分析に大幅な時間がかかったり、分析した結果新たな発見が何も得られないなどのリスクがあるため、本ガイドラインでは対象外としています。

本ガイドラインで扱うデータ利活用の類型

可視化

数値やテキストなどのデータをグラフ、チャート、マップなどの視覚的な形式で表現するプロセス

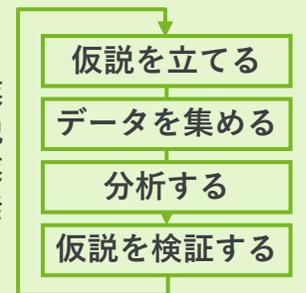


分析

問題事象の原因や傾向など、膨大なデータを探ることで業務に生きる洞察や要点を掴み、課題解決に繋げるプロセス

ワンポイント

分析を行う際は、すぐにデータを収集し分析作業を行うのではなく、まずは仮説を立て、その仮説を検証するために必要なデータの収集・分析作業を行いましょう。これにより、比較的短時間で無駄の少ない分析を行うことができます。※2



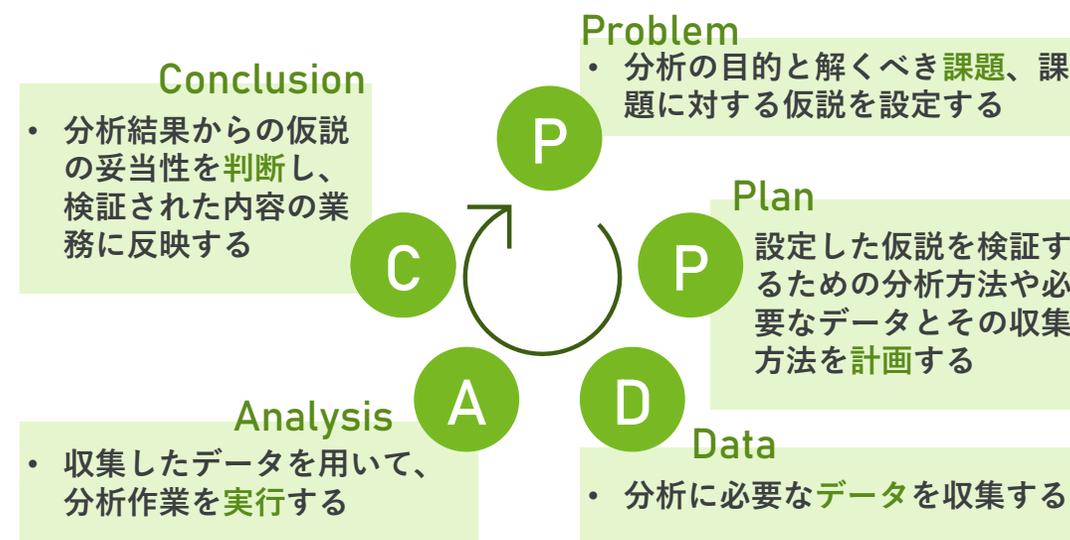
2 ガイドラインの構成

2. 分析で活用できるフレームワーク PPDAC

データ利活用における分析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を活用したデータ分析を前提として説明します。

PPDACのフレームワーク



※1：検証型アプローチの一つである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。

2 ガイドラインの構成

3. データ利活用が業務にもたらす効果

庁内の業務におけるデータ利活用は、以下のような様々なメリットをもたらします。

◆ ニーズに合った施策の検討・実行

データに基づいた状況整理・判断を行うことで、定量的に課題や事象を把握でき、ニーズに合った施策を検討・実行することができます。

また、施策の効果や影響について、データを用いて施策実施前後の状況を比較するなど、定量的な評価を行うことができます。これにより都民に対して、施策の内容や効果について定性的な説明だけでなく、具体的な数値や指標などの根拠に基づいた説明ができるようになります。

◆ 課題の見える化による業務の効率化

データを利活用して複雑な情報を整理することで、現状の課題を発見しやすくなるほか、他職員とも共通の理解を持ちスムーズにコミュニケーション出来るようになります。これにより対応策や改善点を迅速に検討・合意できるという点で、業務効率改善に繋がります。

◆ データ公開による新サービスの創出

各職員が業務でデータを利活用することで、組織内のデータが整備され、データ公開のための準備が整っていきます。行政データを公開することによって、企業や学術研究機関等がそれらを活用できるようになり、都民のQOL向上に寄与する新しいサービスの創出等に繋がります。

2 ガイドラインの構成

4. ガイドライン全体の構成

本ガイドラインは「はじめに」編のほかに、3つの章が存在しています。

第一章では、業務にて最も実践しやすい身近なデータ利活用の類型として、データを見て情報を得ることに焦点をあてて、データの可視化の概要や進め方について説明しています。

第二章では、データを更に深掘りし課題解決に繋げていく分析について、進め方や注意事項をPPDACの5つのフェーズに沿って説明しています。

第三章では、利用許諾取得やデータ品質管理で活用できるより実用的なチェックリストを紹介しています。

本ガイドラインの構成とトピック

はじめに

- 本ガイドラインの背景及び目的
- ガイドラインの構成

第一章： データを 可視化する

- データの可視化の定義と重要性
- データの可視化の進め方

第二章： データを 分析する

- データ分析の概要と進め方
- 問題を特定し仮説を構築する (Problem)
- 仮説検証のための計画を立てる (Plan)
- データを収集する (Data)
- データを分析する (Analysis)
- 分析結果を考察する (Conclusion)
- データ分析の実践例

第三章： チェック リスト

- 利活用許諾チェックリスト
- データ品質チェックリスト

2 ガイドラインの構成

5. 関連する手順書・マニュアルなど

本ガイドラインでは、データを活用するための基礎的な事項を整理していますが、都には他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」で整理している通りです。

例えば、デジタルサービス局のデータ活用事業では、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。

東京都オープンデータ推進庁内ガイドライン※1

東京都におけるオープンデータの推進に向けた基本的な考え方及び取組の方向性について示したガイドライン

データ整備事業データ入力マニュアル※1

「自治体標準オープンデータセット」のレイアウト形式に沿ったデータの作成方法及び更新方法を示したマニュアル

庁内データ連携基盤のデータ活用ガイドライン※1

デジタルツイン庁内データ連携基盤へのデータ掲載にあたり、東京都職員や地理空間データの変換・加工を実施する事業者に対して、データの変換・加工及び利用に関する留意点を整理したガイドライン

※1：各手順書・マニュアルについては、東京都のホームページ（巻末付録：本文で紹介している東京都のサービス・ツール）を参照してください。

第一章： データを可視化する

第一章では、データの可視化の概要や進め方を紹介します。データの可視化は、情報を直感的に理解しやすくするための手段であり、データ活用を行う上での重要なファーストステップです。

可視化の事例などを学びながら、具体的な進め方を理解しましょう。

- 1 データの可視化の概要 P.14
- 2 データの可視化の進め方 P.16

可視化とは、数値やテキストなどのデータをグラフ、チャート、マップなどの視覚的な形式で表現するプロセス

日々の業務において数値やテキストなど様々なデータを扱う際、そのままのデータ形式では情報が複雑・多量で特徴を掴みにくく、全体像の把握が困難です。目的に沿ってデータを可視化することで、複雑なデータセットの中からパターンや傾向、異常などを直感的に把握・理解することが可能になります。データ利活用のファーストステップとして、積極的にデータを可視化していきましょう。

データの可視化の事例

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。



可視化の重要性とメリット

P.15

- データの全体像の直感的把握による課題抽出の簡易化
- 情報共有の円滑化や資料作成時間の短縮による業務の効率化
- データ公開による行政としての透明性の向上

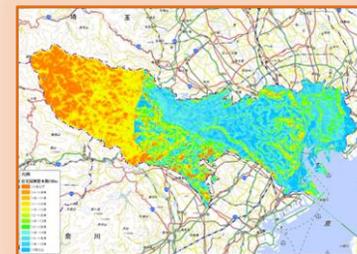
可視化の進め方

P.16

1. 可視化の目的の整理及び要件・手段の確認
2. 必要なデータの収集及び整備
3. グラフやチャートなどの作成

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。



1 データの可視化の概要

1. データの可視化の重要性

庁内の業務において、データの可視化は非常に重要な役割を果たします。

◆ データの全体像の視覚的把握による課題の抽出

データ分析は効果的な施策の検討や意思決定の際に役立ちますが、その前段階として現状を正確に把握したい時には、まずデータを可視化することが有効です。データをそのままの形式ではなくグラフやマップなどを用いて表現することで、データの中に潜む重要なパターンや傾向を視覚的に把握することができ、起きている事象や課題を理解しやすくなります。

例

財政データや都民からの意見を、分野別にグラフで可視化

- ➡ 課題や問い合わせが集中している分野を明確化
- ➡ 具体的な課題抽出・優先的な対応事項を特定

◆ 業務の効率化

グラフやチャートなどを用いて可視化を行うことで、庁内における提案や説明の説得力が増すほか、情報共有や認識合わせが円滑になります。また、ダッシュボードなどを用いてデータの可視化を自動で行うことにより、資料作成などの手間が短縮され、業務効率の向上に大きな効果をもたらします。

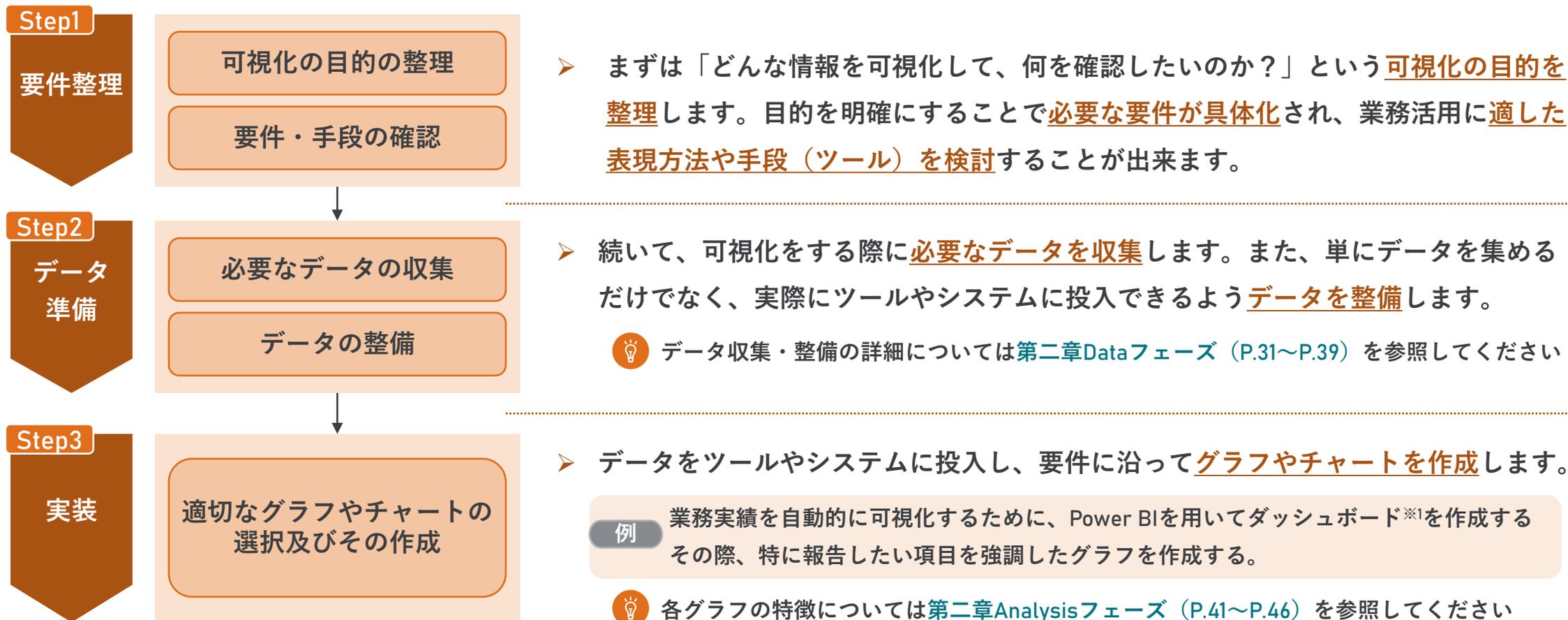
◆ データ公開による信頼性向上

業務で作成したダッシュボードやマップなどを都民向けに公開することで、行政の透明性が高まり、信頼性の向上に繋がります。

2 データの可視化の進め方

1. データの可視化の進め方

可視化の進め方には様々な考え方がありますが、本ガイドラインでは以下の3つのステップで説明します。



※1：ダッシュボードのより詳細な作成方法については、デジタル庁の「ダッシュボードデザインの実践ガイドブック」（巻末付録：リンク一覧②）を参照してください。

2 データの可視化の進め方

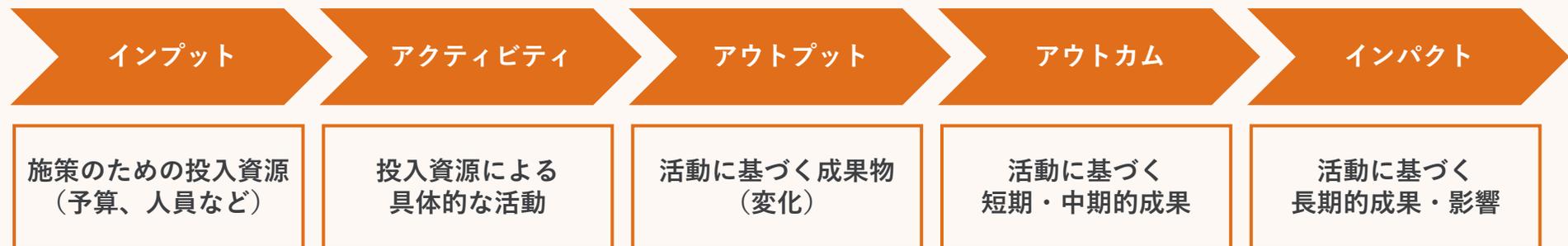
COLUMN ロジックモデルを活用したダッシュボード (1/2)

可視化の目的や要件を整理する際は、様々な事例を参考にしていきましょう。ここでは施策効果を可視化する事例として、ロジックモデルを活用したダッシュボードを紹介します。

近年、施策の成果を明確にする手段として、ロジックモデルという考え方が注目されています。

ロジックモデルとは、事業や施策の計画、実施、評価を体系的に整理し、視覚的に示すフレームワークです。

具体的には、**インプット（資源）・アクティビティ（活動）・アウトプット（成果物）・アウトカム（短期・中期的成果）・インパクト（長期的成果）**の一連の因果関係を図示し、施策の論理をわかりやすく明確に整理することができます。ロジックモデルを用いることで、施策の道筋について関係者間で共通理解を促進できるほか、効果的に施策の状況を把握し評価することが可能になります。



2 データの可視化の進め方

💡 COLUMN ロジックモデルを活用したダッシュボード (2/2)

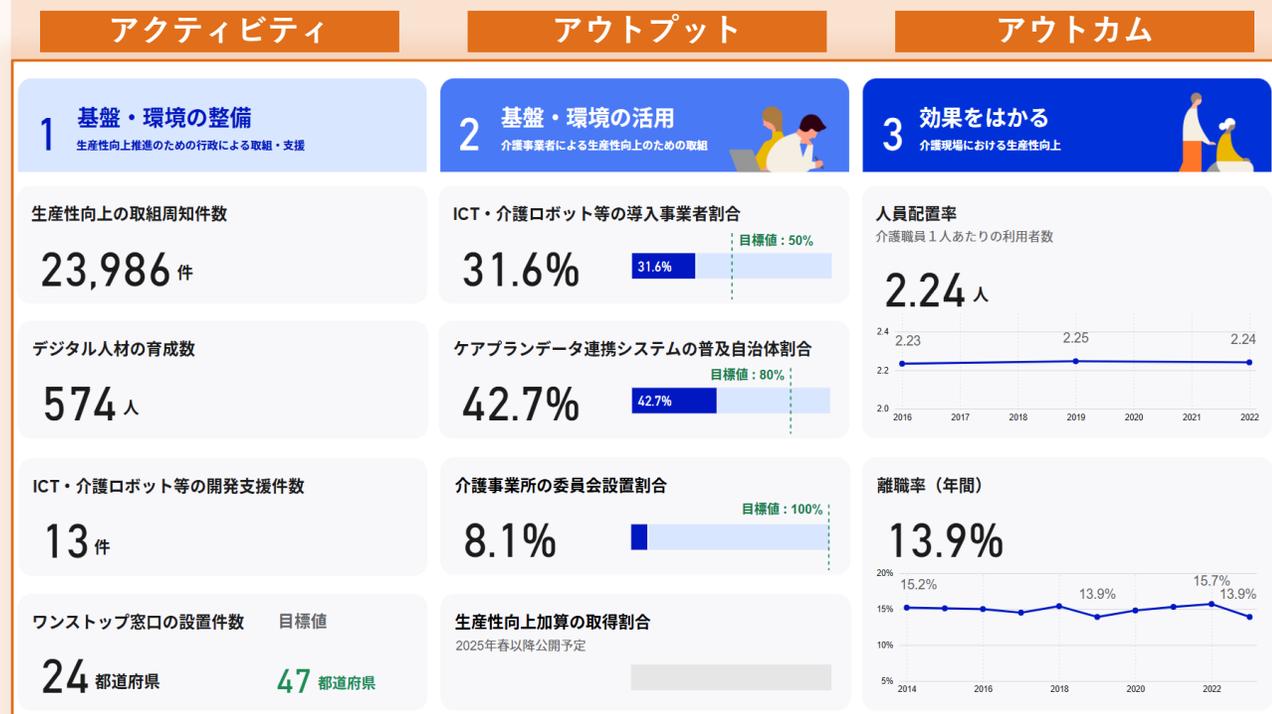
ロジックモデルを活用したダッシュボードとは、ロジックモデルで定義されている各項目の因果関係を視覚的に表現したダッシュボードです。取組内容（アクティビティ）、実績（アウトプット）、効果（アウトカム）などを一つのダッシュボード上に掲載し、プロジェクトの進捗状況や成果を一目で把握することが可能です。

このようなダッシュボードを作成する際には、まず施策の目的から具体的な手段に至るまでの論理的なつながり（ロジック）を明確にします。次に成果（アウトプットやアウトカム）を測るための指標を設定し、ロジックの流れに沿って指標データなどの数値情報をダッシュボード上に配置します。

※1：ロジックモデルのより詳細な作成方法については、文部科学省の「「ロジックモデル」作成マニュアル」（巻末付録：リンク一覧③）を参照してください。

ロジックモデルを活用したダッシュボードを作成し運用していくことで、推進中の事業や施策の効果が明確になり、効果的な意思決定と透明性の向上に繋がります。ぜひ、作成を検討してみてください。

例 介護現場の生産性向上のための主要指（厚生労働省/デジタル庁）



出所：介護現場の生産性向上に関するダッシュボード（デジタル庁）

第二章： データを分析する

第二章では、PPDACの5つのフェーズに沿って、データ分析の進め方や注意事項を説明します。

日々の業務において、どのような場面でデータ分析を実施できるか、データ分析を行うことでどのような効果が期待できそうか、具体的にイメージしながら読み進めていきましょう。

1	データ分析の概要と進め方	—————	P.20
2	問題を特定し仮説を構築する (Problem)	—————	P.23
3	仮説検証のための計画を立てる (Plan)	—————	P.27
4	データを収集する (Data)	—————	P.31
5	データを分析する (Analysis)	—————	P.40
6	分析結果を考察する (Conclusion)	—————	P.51
7	データ分析の実践例	—————	P.55

1 データ分析の概要と進め方

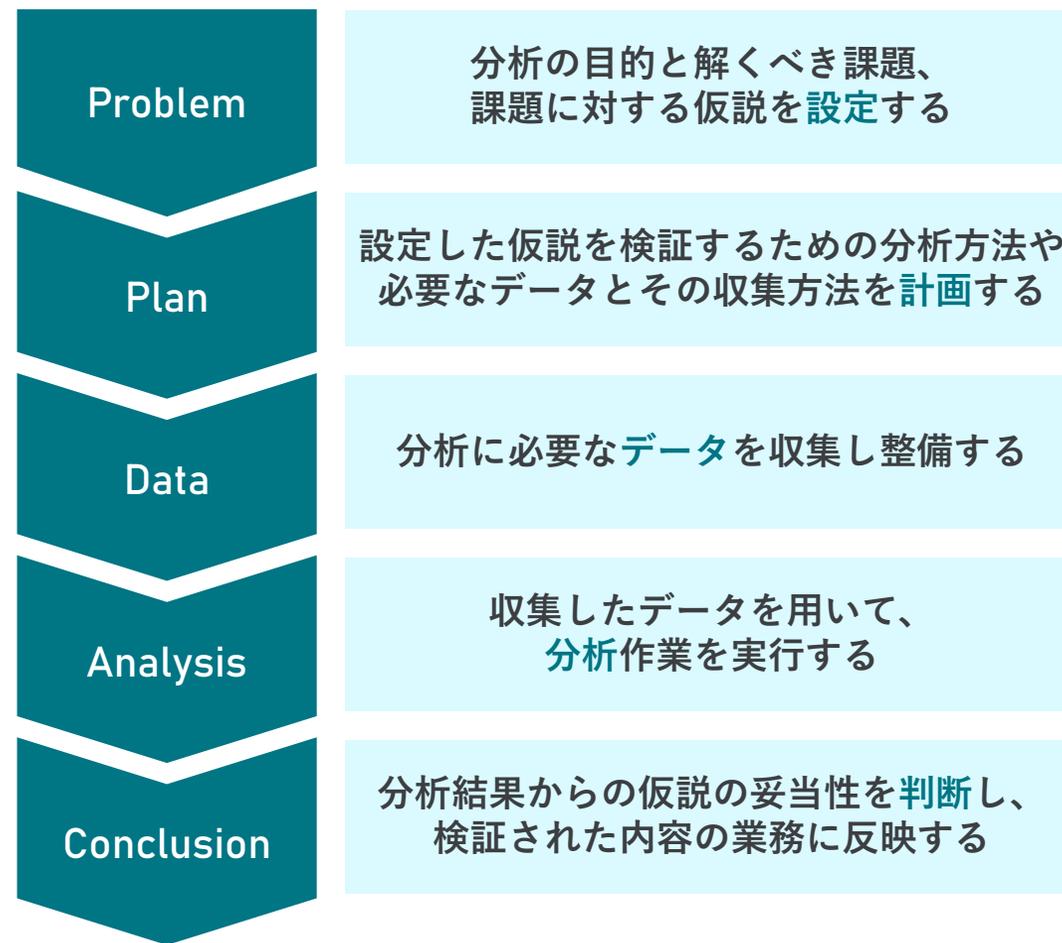
1. 分析で活用できるフレームワーク PPDAC

分析とは、データを整理して起きている事象を把握し、その事象の原因や傾向などを明らかにして課題解決に繋げるためのプロセスです。業務において様々な意思決定や施策検討を行う際に役立ちます。

本ガイドラインでは、データ分析を行う際に「PPDAC」というフレームワークを活用することを推奨しています。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。

本章では、PPDACの各フェーズにおける重要事項や具体的な作業について説明しますが、まずはじめにPPDACの作業を進めるうえで役立つ「PPDACワークシート」というツールを紹介します。

PPDACフレームワーク



1 データ分析の概要と進め方

2. PPDACワークシートの構成

PPDACワークシートは、PPDACの各フェーズ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つから構成されます。各記入箇所を埋めながら作業を進めることで、必要事項の検討漏れや手戻りを予防することができます。

1 分析の目的、解くべき問題を整理

データ分析に取り組む背景や、データ分析を通して解決したい問題と検証したい仮説を記入します。

2 分析で使用するデータを確認

分析に必要なデータやその収集方法を記入します。

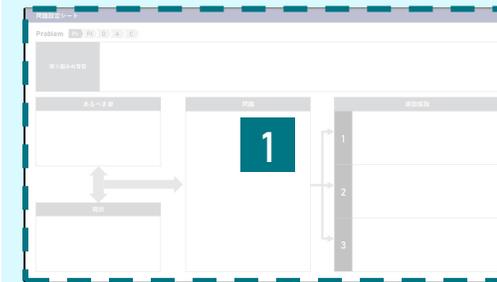
3 分析手法を検討し、分析結果を整理

用いる分析手法や得られた分析結果を記入します。

4 結論と今後の施策の方向性を整理

分析結果を踏まえた結論と、今後の施策の方向性を記入します。

PPDACワークシートの構成



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。

...各項目の具体的な記入方法・記入ステップはPPDAC各フェーズの中で説明しています

1 データ分析の概要と進め方

3. 分析の進め方とワークシートの活用方法

分析の進め方には様々な考え方がありますが、本ガイドラインでは以下の3つのステップで説明します。

Step1

分析の目的・
解くべき課題を整理
(Problem)

- 分析を始める際は、まず分析の目的や解くべき課題を整理します。問題設定シート（ワークシート1枚目）を活用して、分析の背景や目的、解くべき課題とその課題の原因として考えられる仮説を明確にしましょう。

問題設定シート

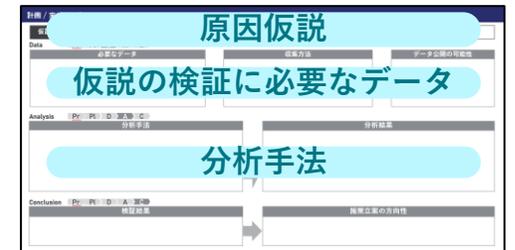


Step2

分析計画を立てる
(Plan)

- 続いて、原因仮説を検証するために必要なデータや分析手法を検討し、分析計画を立てます。原因仮説ごとに計画/実行シート（ワークシート2枚目）を用いて、分析の計画を整理しましょう。

計画/実行シート

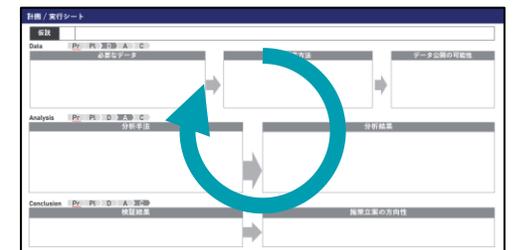


Step3

分析を実行する
(Data/Analysis
/Conclusion)

- 最後に、実際にデータを収集し分析作業を行います。このとき、分析作業の結果に基づいて、計画/実行シート（ワークシート2枚目）に記載していた内容を更新しましょう。

計画/実行シート



...各ステップにおける具体的なワークシート記入方法はPPDAC各フェーズの中で説明しています

データ分析によって解決すべき問題を設定するフェーズ

データ分析を行う際は、作業を始める前に分析によって解決すべき問題を明確にすることが重要であり、その問題設定を行うのがこのProblemフェーズです。ここで分析の目的を明確に定めることで、後続で適切にデータ分析の計画を立て、実践に繋げていくことが出来ます。

PPDACワークシート (P.26で詳細説明)

Problemフェーズでは、
解くべき問題を整理し原因仮説を記入します

あるべき姿・現状

問題

あるべき姿と現状を具体的・定量的に
記入し、解決すべき問題を特定します

原因仮説

その問題がなぜ発生しているのか、原因を検討し仮説を立てます
(ワークシート内には最大3つの仮説を記入することが可能です)

問題設定シート

必ず守るポイント

- Problemフェーズは、「問題の明確化」と「原因仮説の構築」の2つのステップで進めましょう
- 「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事象を具体化しましょう P.24
- 「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しょう P.25

重要な考え方

- 問題は、あるべき姿・現状を具体的かつ定量的に整理することで、解像度高く明確化することが出来ます
- 原因仮説を構築する際は、原因を幅広く論理的に考えることと、「その問題が生じたのはなぜか？」を繰り返しながら深く掘り下げて思考することが重要です

2 問題を特定し仮説を構築する (Problem)

1. 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。

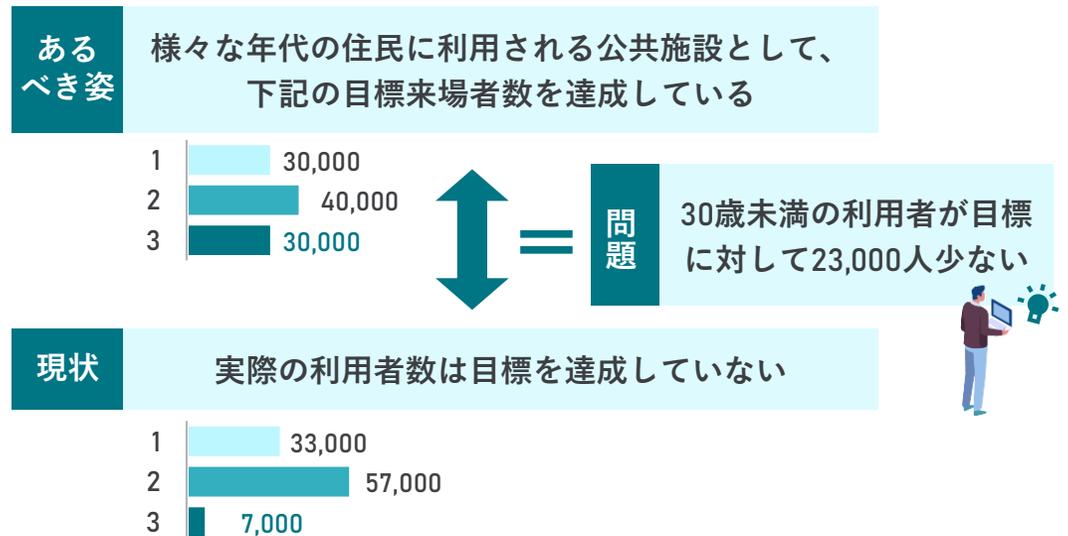
「利用者が少ない」のような曖昧な問題を設定してしまうと、分析計画を立てることが難しくなってしまいます。

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

問題 NG例
利用者が少ない

問題が曖昧で何を分析すれば良いかわからない…

ワンポイント

定量的かつ具体的な問題の設定は、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繋がります。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。

2. 原因仮説の構築

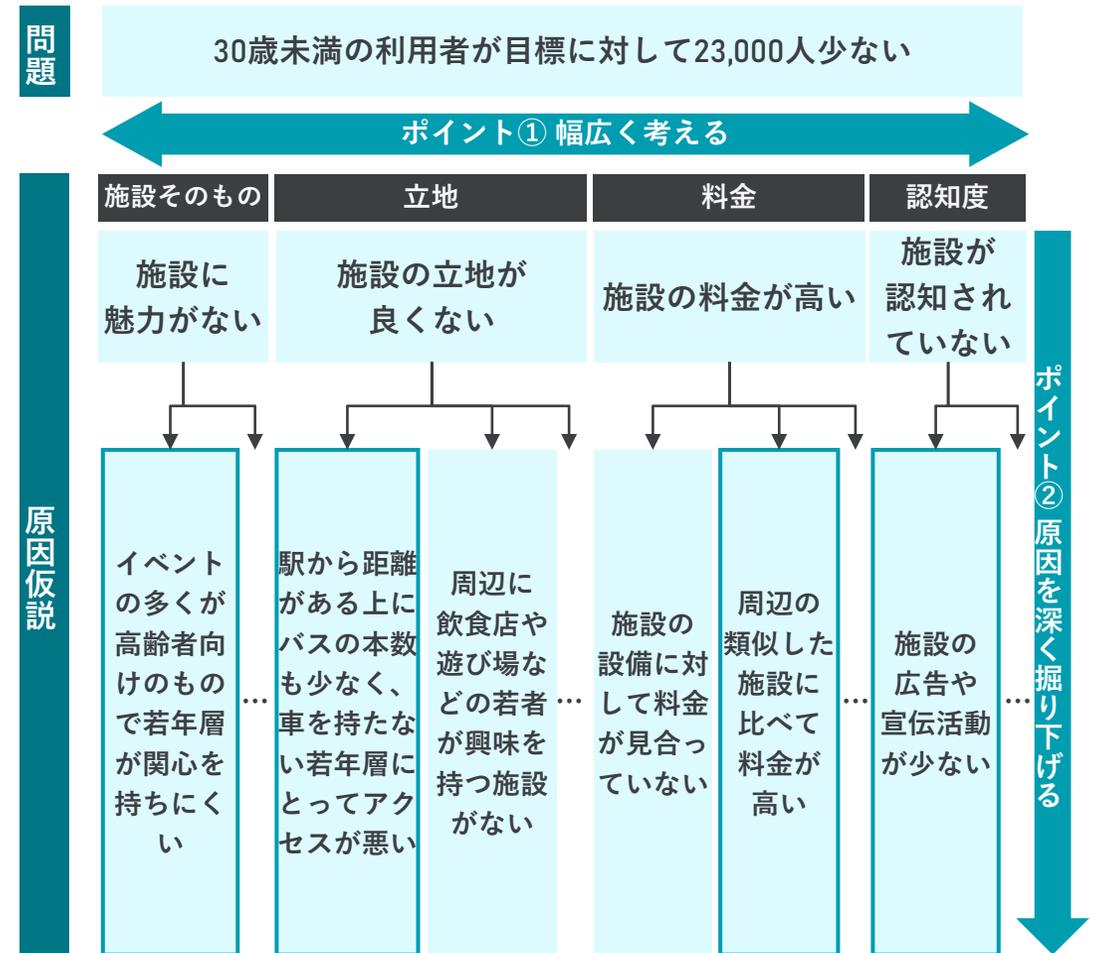
「原因仮説の構築」では、問題の主たる原因に関する仮説※1を立てます。ここでは次の2点が重要です。

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原因を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、その中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。選んだ原因仮説の妥当性はこの後のフェーズで検証していくため、この時点では正確な根拠は無くても大丈夫です。

「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

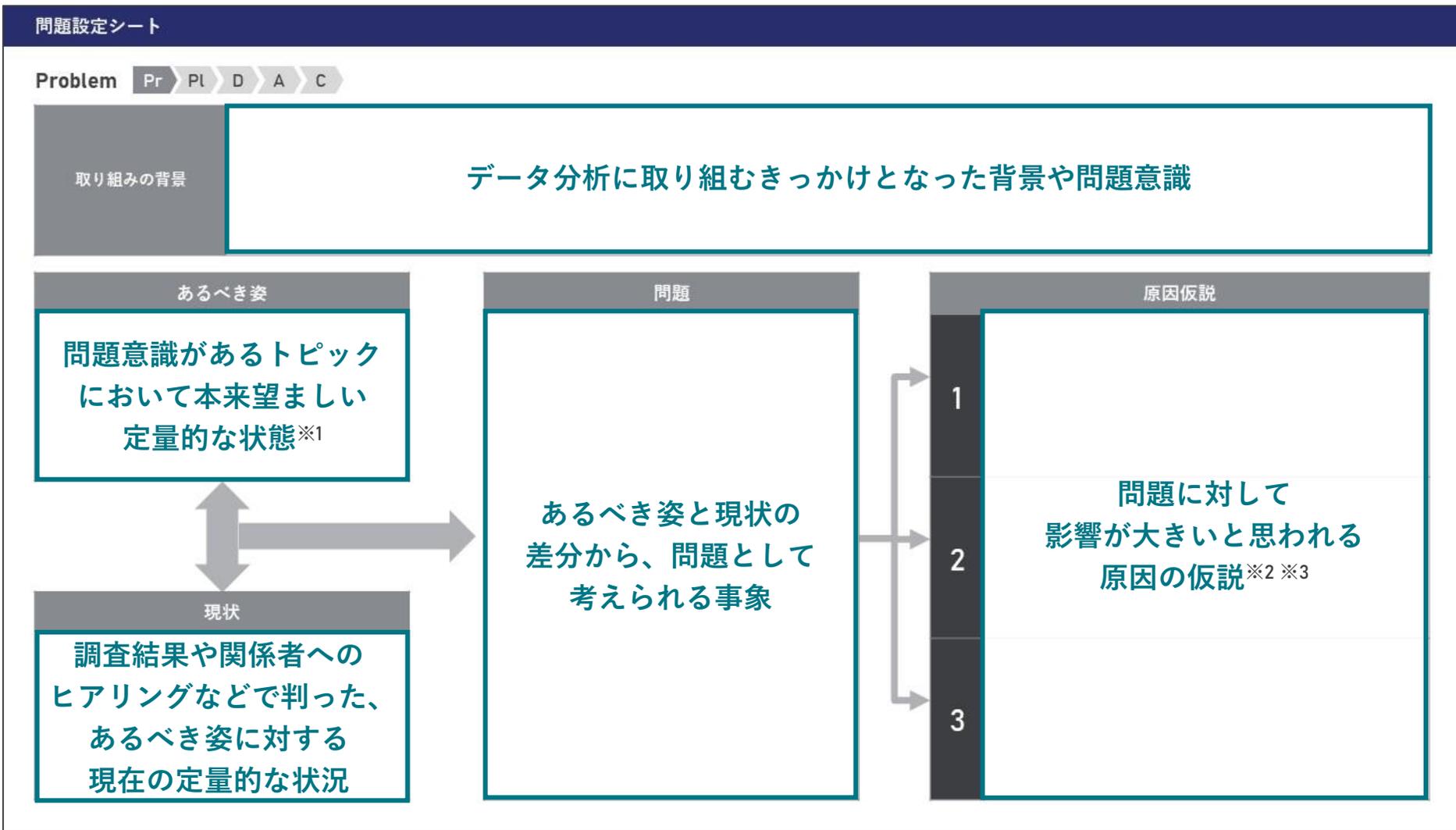
※1: 仮説とは、その時点の情報から導くことができる仮の答えを意味します。仮説を立てることで検証すべきことが絞られ、効率的に分析を進めることができます。この仮説を分析によって検証していくこととなります。

※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。

2 問題を特定し仮説を構築する (Problem)

3. Problemフェーズにおけるワークシートの使用法

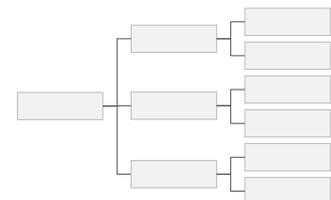
Problemフェーズでは問題設定シートの各項目を使用して、分析の前提となる背景・仮説を整理します。



※1: シン・トセイや各部署・事業の目標も参考にすると良いです。

※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ありません。

※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します。詳しくは、P.25を参照ください。



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るフェーズ

Planフェーズでは、Problemフェーズで整理した原因仮説を検証するために必要なデータや分析方法を検討して、データ分析の計画を作成します。データ収集などの時間を要する作業へ着手する前に計画することで、目的と合致しない方法で分析してしまう等のリスクを抑止できるほか、作業全体を俯瞰してその妥当性を高めることができます。

🔗 PPDACワークシート (P.29で詳細説明)

Planフェーズでは、分析計画の作成を行います

必要なデータ

収集方法

データ公開の可能性

分析に必要なと考えられるデータと、その収集方法を記入しましょう。またデータ公開をするかどうかについても記入しましょう

分析手法

分析結果

分析の際に使用する手法を検討し、想定される分析結果を記入しましょう

計画/実行シート

🎯 必ず守るポイント

- 計画を作成する際は、まずどのようなデータを用いてどのような分析を行うべきかを整理しましょう。また、データをどのように収集するか、についても検討しましょう
- 「PPDACワークシート」を活用しましょう。
このシートを使用することで、分析着手前に各フェーズでの作業内容を俯瞰しながら計画を立てることができ、検討漏れや手戻りを予防できます

P.28

🎯 重要な考え方

- 分析手法は、検証結果として欲しい情報がわかりやすく表現される手法を、優先的に選択します
- 💡 分析手法の詳細は、P.41~P.46 (Analysisフェーズ) を参照してください
- データの収集方法は、事業のスケジュールや予算に見合った方法を優先的に検討します

P.28

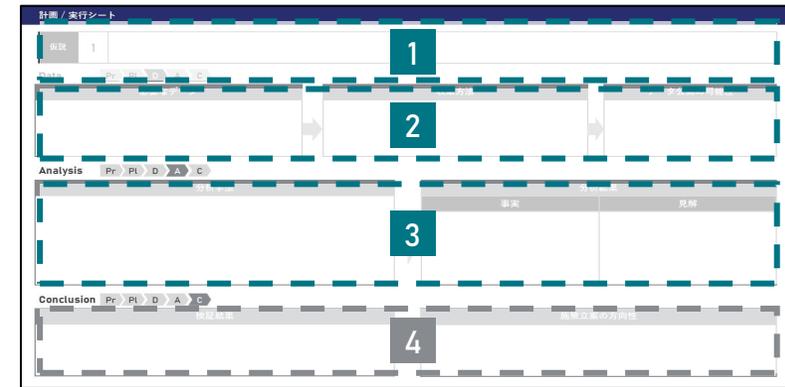
3 仮説検証のための計画を立てる (Plan)

1. Planフェーズにおけるワークシートのご使用方法

Planフェーズでは、Problemフェーズで立てた仮説を検証するために必要なデータやその収集方法、分析手法を検討し、分析結果を予想します。データの収集方法については、事業のスケジュールや予算に見合った方法を検討しましょう。また、分析手法については、検証結果として欲しい情報がわかりやすく表現される手法を選択しましょう。

Planフェーズで分析計画を立てる際は、計画/実行シートを使用しましょう。実際の分析作業に着手する前に計画を立てることで、全体の作業内容を確認し、検討漏れや手戻りを予防することができます。なお、計画/実行シートはまずPlanフェーズ時点での想定を記入（Data/Analysisフェーズ欄）しますが、後続のフェーズを進める中でPlanフェーズ時に記載したData/Analysisフェーズ欄を更新し、Conclusionフェーズ欄を新たに記載します。

計画/実行シート（ワークシート2枚目）



ワークシート記入箇所

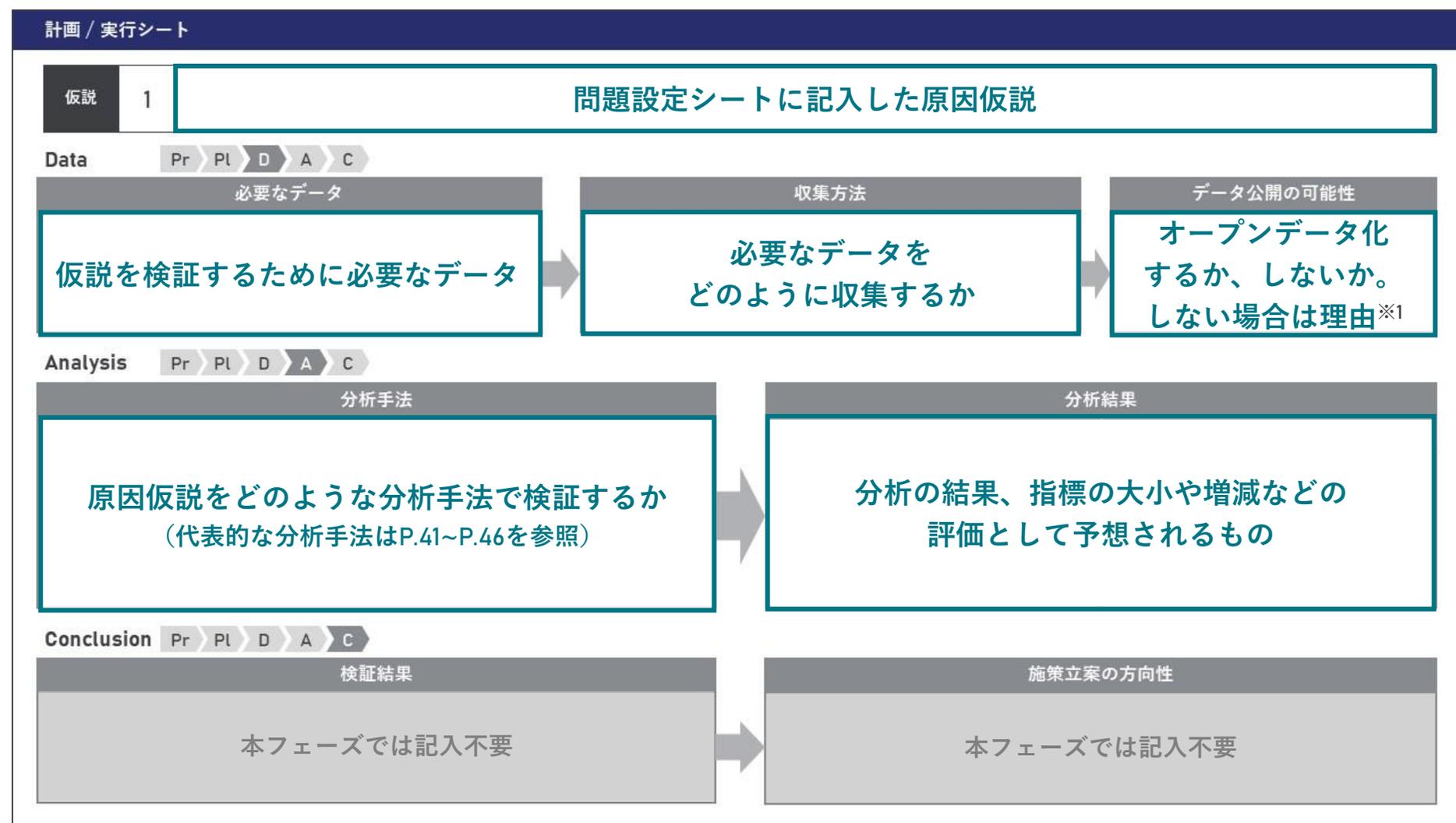
- | | |
|---------------------------|---------------------------------------------------------------------|
| Plan
フェーズ | 1 Problemフェーズで立てた原因仮説のうち、検証する仮説を一つ選択し記入 |
| Data・
Analysis
フェーズ | 2 分析に必要なデータやその収集方法の想定をDataフェーズ欄に記入 |
| Conclusion
フェーズ | 3 用いる分析手法や分析結果の想定をAnalysisフェーズ欄に記入 |
| | 2 3 Planフェーズ実施後に、実際に収集したデータや分析結果に応じてData/Analysisフェーズ欄に記載されている内容を更新 |
| | 4 Planフェーズ実施後に、仮説の検証結果をConclusionフェーズ欄に記入 |



3 仮説検証のための計画を立てる (Plan)

1. Planフェーズにおけるワークシートでの使用方法

Planフェーズでは原因仮説に対して計画/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。



※1: 委託調査やアンケート調査などで、都がデータを独自に収集する場合は、オープンデータ化の要否を検討します。
 なお、国及び地方公共団体は、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ます。

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する必要があります。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。

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>

※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④) 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。

Planフェーズで立てた計画をもとに、必要なデータを集めて整備するフェーズ

Dataフェーズでは、データの収集・整備を行います。データを集める際の注意点やその取扱い、データを整備する際の留意事項を確認しましょう。

また、収集したデータをオープンデータとして公開することも重要です。データ公開に関連した取組についても意識しましょう。

PPDACワークシート（P.39で詳細説明）

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

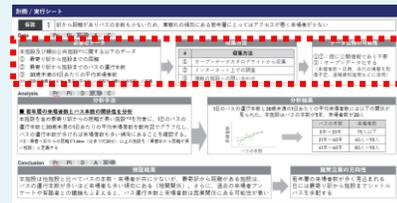
必要なデータ

収集方法

実際に収集できたデータと、その際の収集方法を上書きしましょう

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



必ず守るポイント

- データを集める際は、そのデータを使う権限が都にあることを必ず確認しましょう。また、組織として取得したデータを安全に管理する仕組みを整えましょう
- 分析作業をスムーズに進めるために、データを集めるだけでなく「分析しやすい状態」に整備しましょう

重要な考え方

- ✓ データ品質の評価観点は多くありますが、都では正確性・完全性・一貫性・可用性・最新性の5つを重視しています
- ✓ 都や自治体全体での標準化の観点から、収集したデータはその項目や様式が一般的なものか意識します
- ✓ 計画通りのデータ収集が難しい場合は、他の収集方法の検討や分析方法の変更、類似データでの代用などで対応できます

💡 データ品質を確認する際は第三章のチェックリストを活用しましょう

P.37

1. データの利用許諾取得

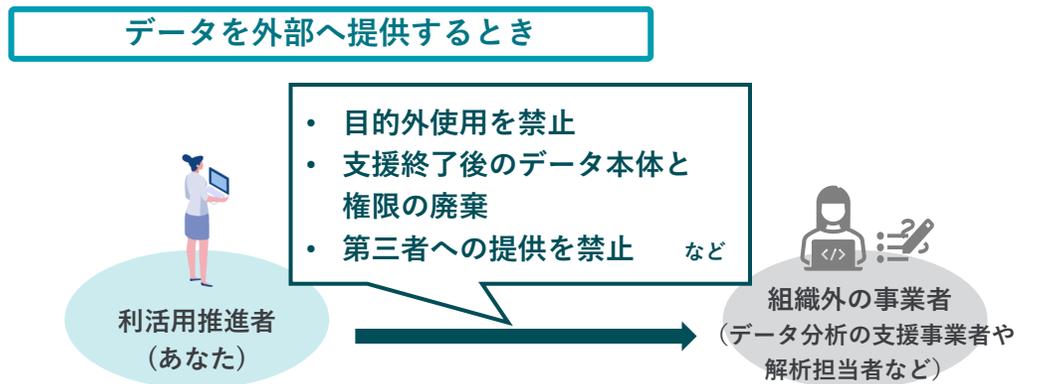
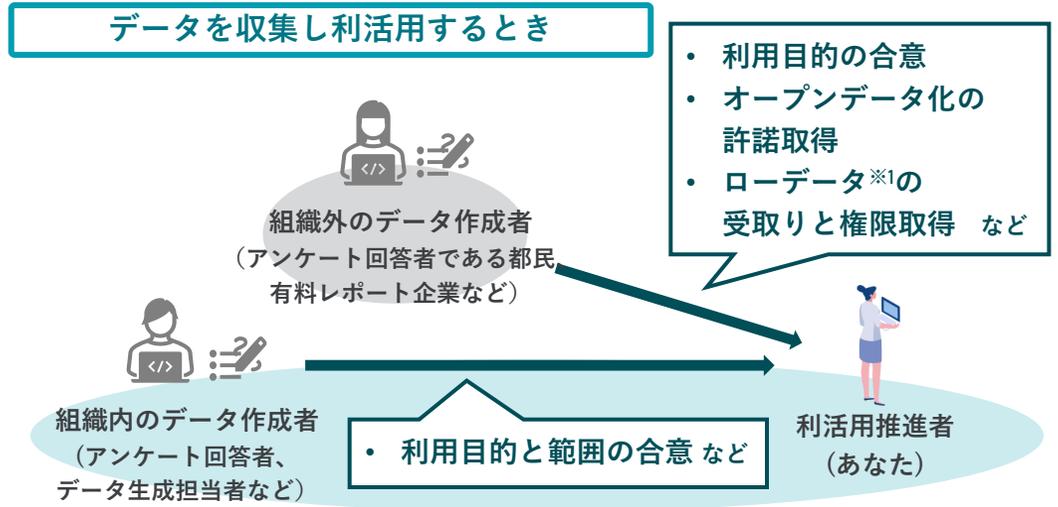
データを収集し利活用する際は、まずデータの利用許諾を確認する必要があります。利用許諾を確認していない場合、法的観点から分析結果を業務で活用できないなどのリスクに繋がる恐れがあるためです。

利用許諾に関する注意事項はデータの収集・利活用の状況に応じて様々です。例えば、組織内でデータを収集する場合はデータ提供者から利用目的などの合意を得る必要があるほか、調査委託など組織外から収集する場合は、一事業や部署に閉じた限定的な利用目的・範囲の合意ではなく、ローデータ※1の受取りやデータの継続的な利用権限なども合意できないか検討しましょう。また、データを組織外に提供する場合は、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。

※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該当します。

※2：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」（巻末付録：リンク一覧⑤）を参照ください。

1 利用許諾の取得のために注意すべき事項※2



4 データを収集する (Data)

フェーズ

Pr

Pl

D

A

C

Data

2. データの安全性管理

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う **安全性管理** の観点も非常に重要です。

この“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を発生させないようにするための取組です。例えば、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、組織内収集データと外部収集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範囲などの権限を確認するよう、ルールとして運用していくことも重要です。

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

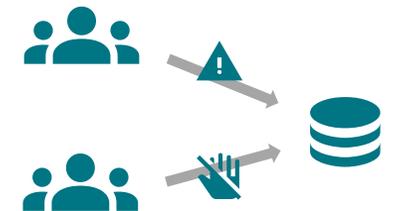
- データアクセスへのパスワード設定と定期的な変更

+

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



自身がデータ利活用を終えた後も、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

4 データを収集する (Data)

フェーズ

Pr

Pl

D

A

C

Data

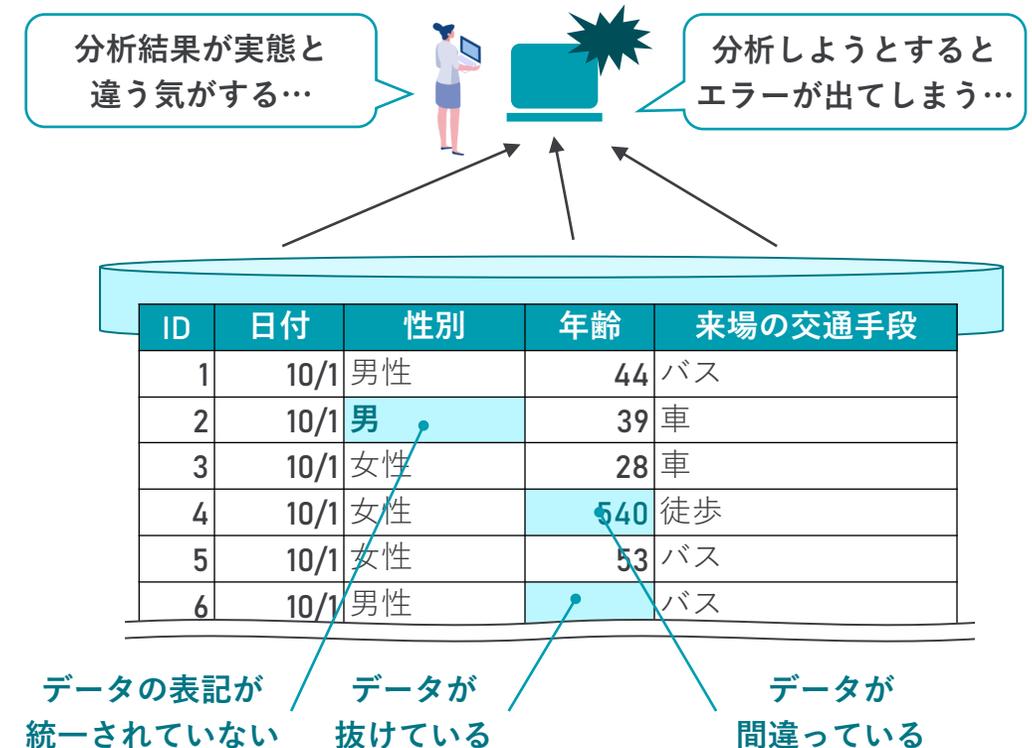
3. データの品質管理 品質管理の必要性

データの品質※1を適切に管理することは、データ利活用における最重要事項の一つです。分析時に利用するデータの品質が低いと、適切な結果が得られなかったり、データ品質の改善作業に膨大な時間をがかかってしまうなどの様々な問題が生じるほか、結果的にデータを活用して提供しているサービスの品質も下がってしまう恐れがあります。サービス品質の向上や個々の作業負荷の軽減にむけて、データを収集し整備する際は適切なデータ品質となるよう心掛けましょう。

ただし、データ品質とデータ粒度を混同しないよう注意が必要です。細かい粒度のデータが必ずしも高品質であるというわけではないうえ、細かいデータの収集・管理にはコストもかかります。データの利用目的に照らして適切な粒度を検討しましょう。

※1：データ品質についてより詳細な情報を確認したい場合は、デジタル庁の「政府相互運用性フレームワーク(GIF) 460 実践ガイドブック」（巻末付録：リンク一覧⑥）を参照してください。

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

4 データを収集する (Data)

フェーズ

Pr

Pl

D

A

C

Data

3. データの品質管理 データ品質の評価

高品質なデータの管理を行うためには、現在のデータ品質を評価したうえで改善に向けた整備方法を検討する必要があります。このとき、データ品質評価は主観的に行うのではなく、客観的に行いましょう。

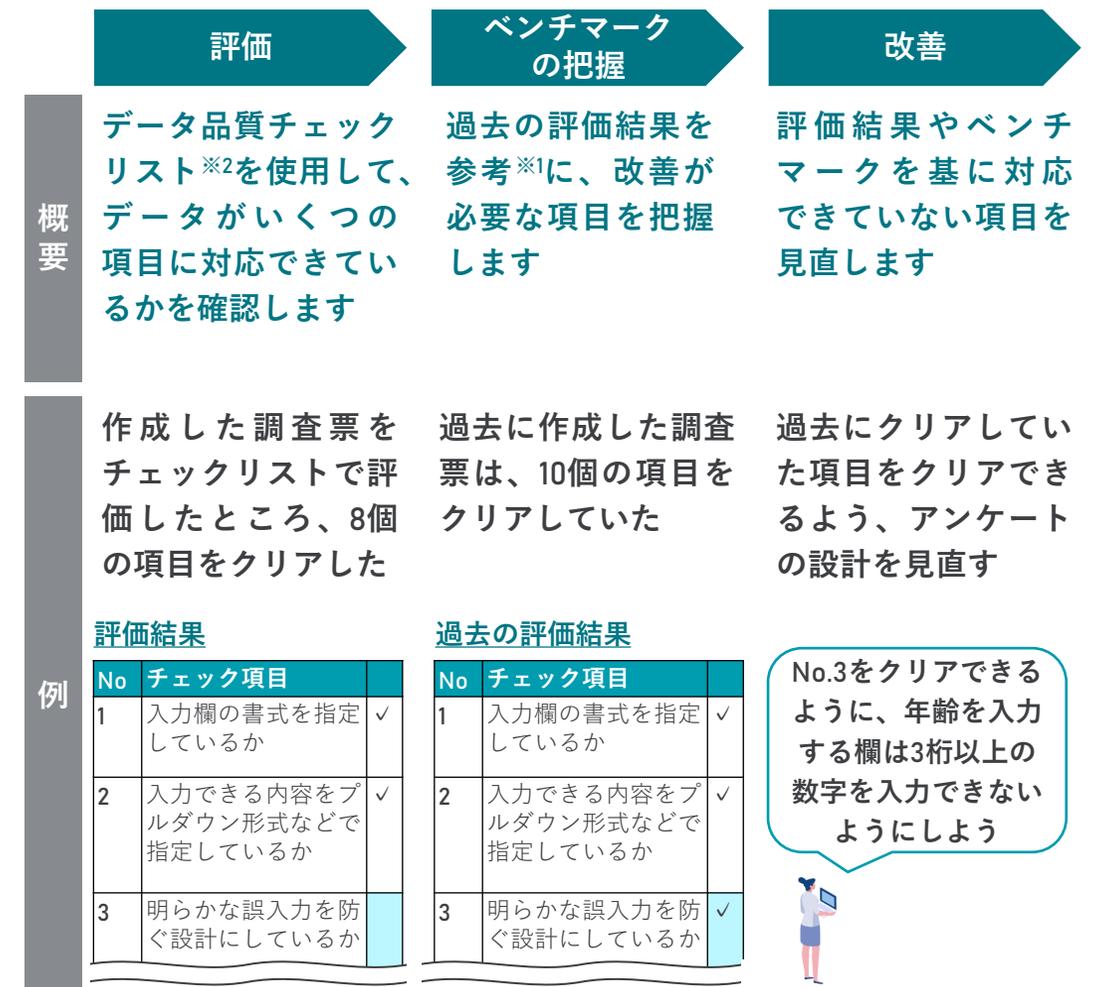
例えば、過去に類似したデータを評価している場合は、評価結果をベンチマーク※1として活用し改善します。ベンチマークを把握することによって、評価結果が良いかどうかを客観的に判断できるようになります。

また、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、「データ品質チェックリスト」※2を使用してデータ品質を評価・改善を行いましょう。

※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第三章「データ品質チェックリスト」を参照ください。

データ品質の改善プロセス



4 データを収集する (Data)

フェーズ

Pr

Pl

D

A

C

Data

3. 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第三章の「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	 <p>必要な時にアクセス可能</p>	 <p>システム停止などによりアクセス不可</p>												
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>2025/4/1</td> </tr> </tbody> </table>	名称	更新日時	アンケート結果.xlsx	2025/4/1	<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>2021/10/1</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	2021/10/1				
名称	更新日時													
アンケート結果.xlsx	2025/4/1													
名称	更新日時													
アンケート結果.xlsx	2021/10/1													

※1：デジタル庁の「政府相互運用性フレームワーク(GIF) 460 実践ガイドブック」（巻末付録：リンク一覧⑥）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。

4 データを収集する (Data)

3. データの品質管理 データの標準化

データを扱う際は「標準化」についても意識しましょう。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す。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。



自治体標準オープンデータセット

「自治体標準オープンデータセット」とは、オープンデータの公開とその利活用促進を目的とし、政府として公開を推奨するデータと、公開するデータの作成にあたり準拠すべきルールやフォーマット等を取りまとめたものです。定義されている主なデータセットは下記の通りです。

自治体標準オープンデータセット

公共施設一覧	教育機関一覧	文化財一覧
公営駐車場一覧	地域・年齢別人口	公営駐輪場一覧
公衆無線LANアクセスポイント一覧	公衆トイレ一覧	
ゴミの分別方法一覧	小中学校通学区域情報	
ゴミ集積場所一覧	子育て施設一覧	赤ちゃんの駅
介護サービス事業所一覧	ボーリング柱状図等	
都市計画基礎調査情報	...	

脚注：「政府相互運用性フレームワーク (GIF)」や「自治体標準オープンデータセット」を活用する際は、デジタル庁のホームページ（巻末付録：リンク一覧⑦・⑧）を参照ください。

COLUMN 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繋がることが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、データの特長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。

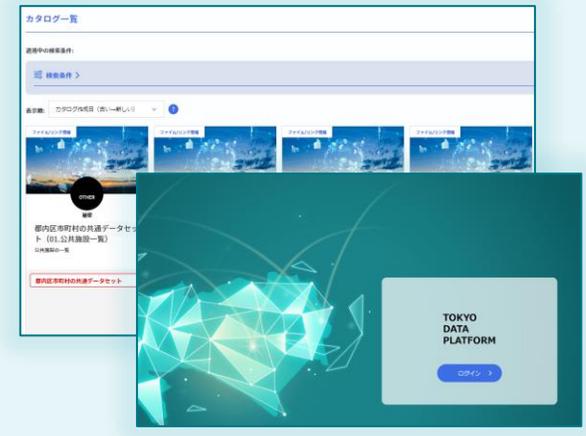
※1: デジタル庁の「オープンデータ基本指針」(巻末付録: リンク一覧⑨)では、オープンデータとは二次利用可能なルールが適用されており、かつ機械判読に適した、無償で利用できるデータと定義されています。

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など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)

はじめに

第一章

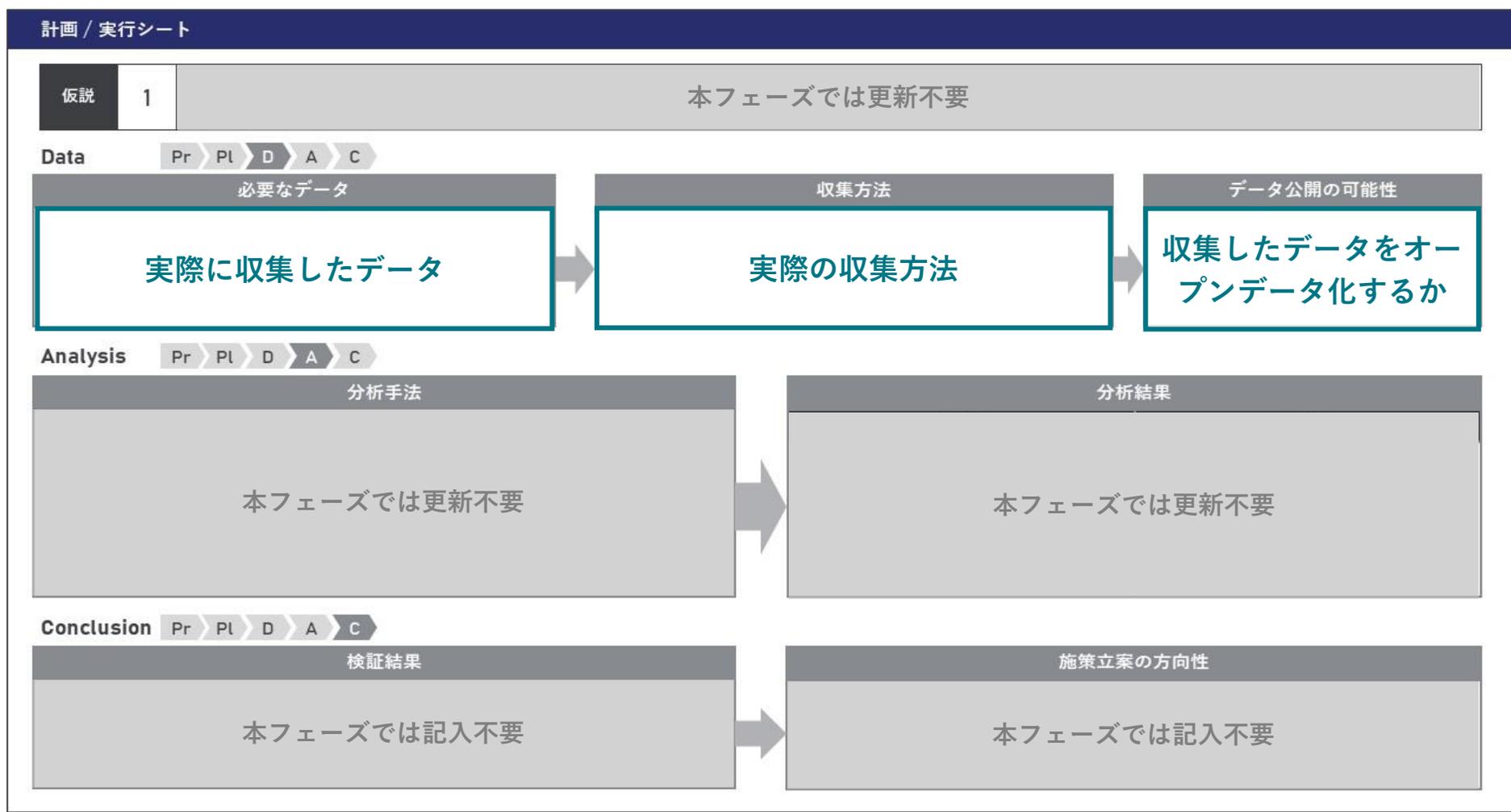
第二章

第三章

4 データを収集する (Data)

3. Dataフェーズにおけるワークシートでの使用方法

Dataフェーズでは、実際に収集したデータやその収集方法をワークシートに記入します。Planフェーズで計画した内容を上書きしましょう。



これまで立てた計画にもとづき実際に分析作業を実施するフェーズ

Analysisフェーズでは実際に分析を行います。

分析作業は、その目的や原因仮説、データやツールなどにより異なります。様々な分析観点の特徴や活用シーン・注意点を学び、自分のデータ分析の目的に沿った分析方法をイメージできるようになりましょう。

PPDACワークシート (P.50で詳細説明)

Planフェーズで記載した内容を、
実際に行った分析手法や分析結果に書き換えます

分析手法

実際に行った分析手法に
更新しましょう



分析結果

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しましょう

必ず守るポイント

- 分析の観点として、違いを見る・変化をさぐる・全体を見わたす・関係をさぐるの4つの主要な観点を目的にあわせて使い分けましょう P.41~P.46
- 施策効果を測る際は、前後比較を行うことが有効です。比較が行えるよう、施策実施後だけでなく施策実施前のデータも事前に収集しておきましょう。また、測定された施策効果に、施策以外の外部要因の影響が含まれていないか注意しましょう P.42

重要な考え方

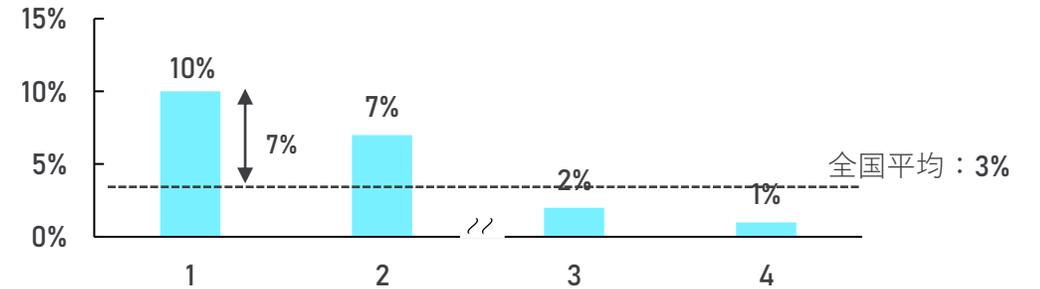
- ✓ データを比較する際は定義や条件を揃えることを意識します P.47
- ✓ 分析の際はデータのサンプル数が十分であり、結果に偏りが出るような収集方法でないことを確認します P.48
- ✓ データに外れ値がある場合は、外れ値が存在する理由を確認し、その値を除外すべきか検討します P.49

1. データを分析する観点①：違いを見る

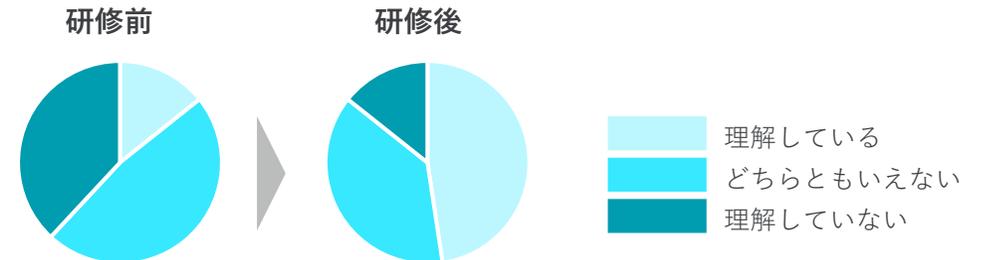
どんな観点？	実数や比率の差から示唆を抽出
主な利用シーン	1つの指標に関して複数項目間の状況や傾向の違いを把握したいときに使用（施策実施の前後比較や数字の大小比較など）
代表的なグラフ	棒グラフや円グラフ
具体例	<p>例1 都道府県別の空き家率の比較 都道府県別の空き家率を示している。A県は都道府県のなかで最も空き家率が高いので、空き家が発生しやすい要因がある</p> <p>例2 研修前後の理解度の比較 研修前後の理解度の比較を示している。研修後は「理解している」の割合が増えたことから、研修を通して理解度が高まったことがわかる</p>

例1 都道府県別の空き家率の比較

グラフ タイトル



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

5 データを分析する (Analysis)

💡 COLUMN 施策実施前後の違いを見る

ある出来事の影響や状況の変化を評価する方法として、**前後比較**という手法があります。これは前後の状態の違いを見て、その差分を特定する方法です。

業務においては、**施策の効果**を測る際に前後比較を活用することが有効です。施策実施前の状態と施策実施後の状態を比較することで、施策の効果を定量的に測ることができます。

施策を実施する際は、前後比較によってその効果が測れるよう、施策実施後だけでなく**施策実施前のデータも事前に収集**しておきましょう。また、前後比較によって施策効果を測定した際は、測定された効果に**施策以外の外部要因（業務繁忙期による変動、他の施策の効果など）の影響**が含まれていないか注意しましょう。

例

業務システム導入前後の業務量の比較
(時期によって業務量が変わる場合)

✔ 前年同月との比較により、**業務繁忙期の影響（外部要因）が除かれている例**

	システム導入前 (2023年9月)	システム導入後 (2024年9月)	導入の効果
ひと月あたりの 作業時間 (時間)	80	50	30

システム導入前のデータの
事前収集が必要

導入前と導入後の差分が
導入の効果となる

⊘ 前年別月との比較により、**業務繁忙期の影響（外部要因）が含まれている例**

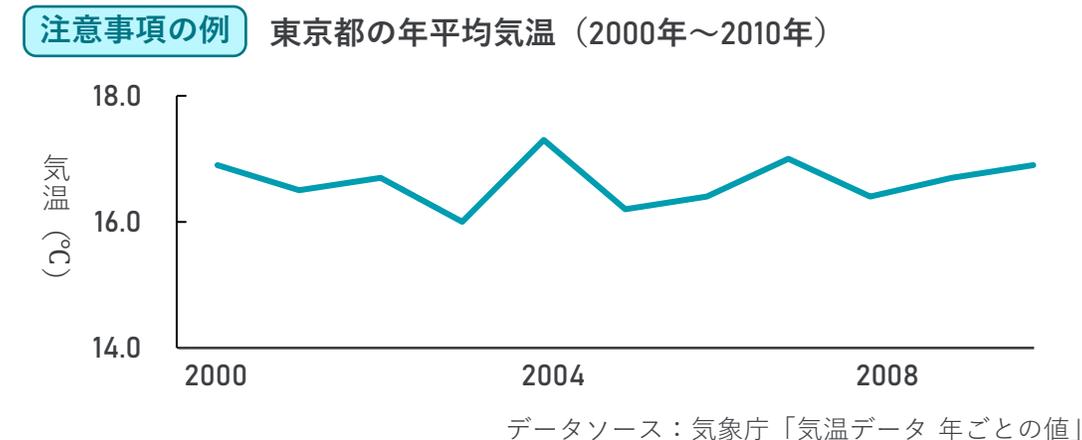
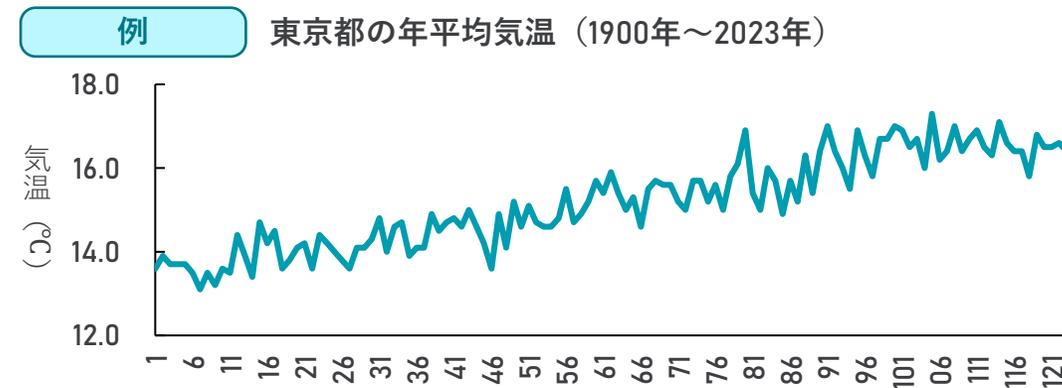
	システム導入前 (2023年9月)	システム導入後 (2024年12月)	導入の効果
ひと月あたりの 作業時間 (時間)	80	70	10

システム導入の効果以外に、外部要因
(時期の違いによる業務量の差) が含
まれていることに注意

5 データを分析する (Analysis)

1. 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>例 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>注意事項の例 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>



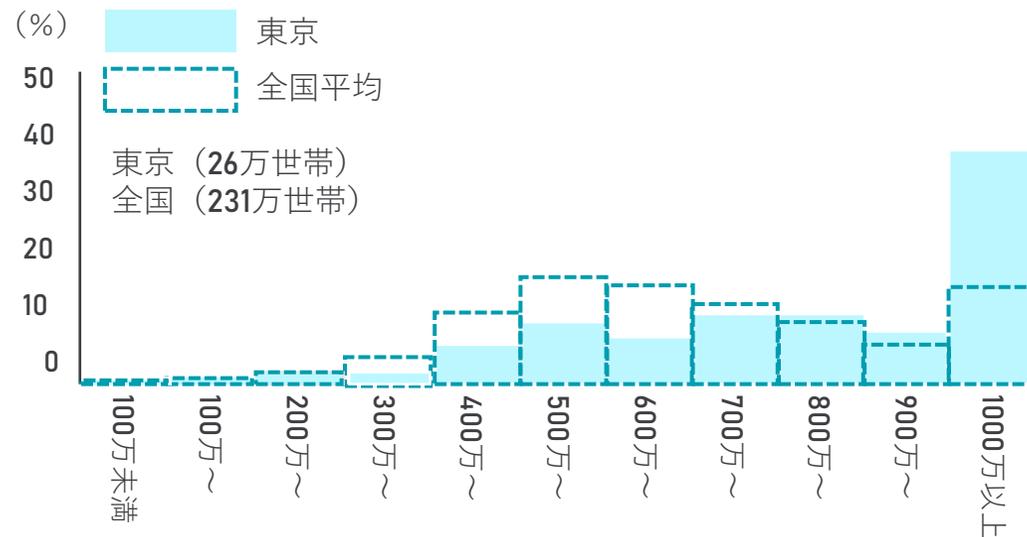
ワンポイント

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

1. 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	<p>例 30代子育て世代の世帯年収分布図</p> <p>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とわかる</p>

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。

1. 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通りです。

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**※1のままであり、クラスの実態をより正確に把握できます。

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、その定義は以下の通りです。

生徒	A	B	C	D	E
点数	20	34	38	44	44
偏差	-16	-2	2	8	8

44点から平均値の36点を引く

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 (標準偏差 = $\sqrt{\sum(\text{各値} - \text{平均値})^2 / \text{サンプル数}}$)	8.9



※1: データの数が偶数の場合は、「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。

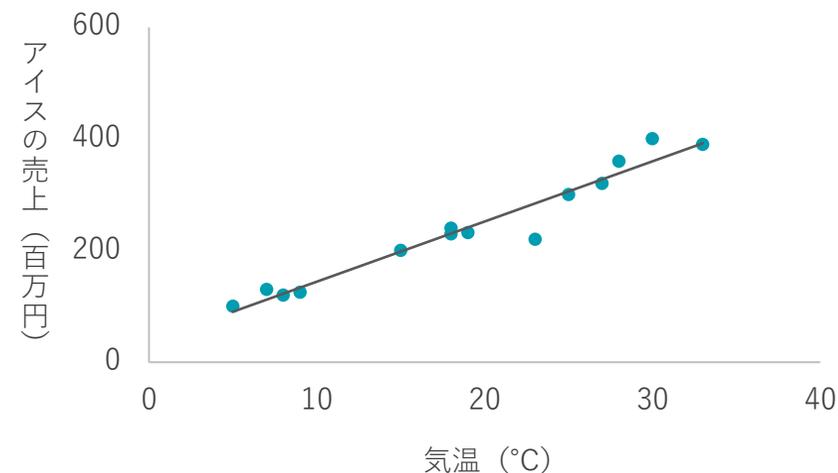
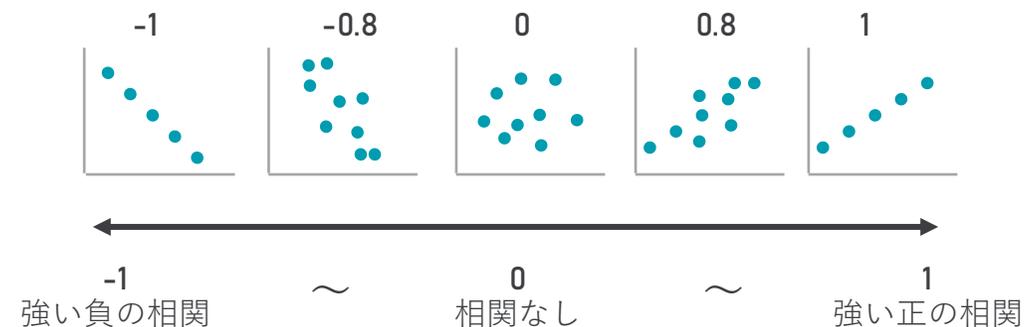
5 データを分析する (Analysis)

1. 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	<p>例 気温とアイスの売上の散布図</p> <p>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</p>


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**の通り、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

5 データを分析する (Analysis)

2. 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繋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

5 データを分析する (Analysis)

フェーズ

Pr

Pl

D

A

C

Analysis

2. 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

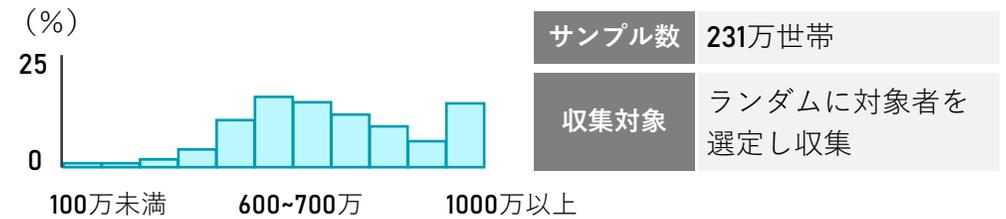
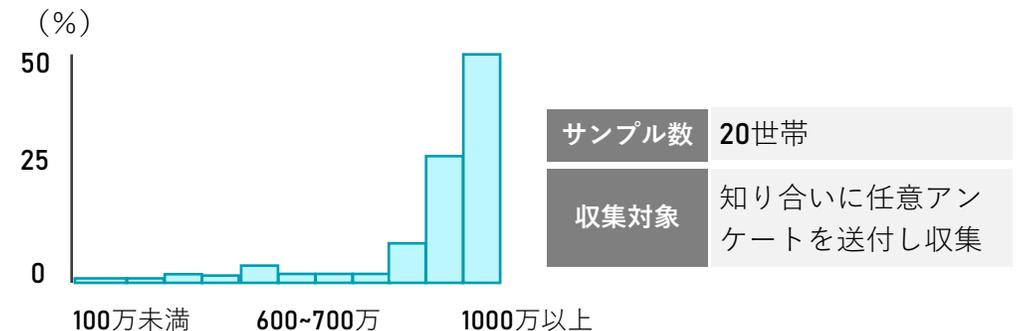


図2

サンプル数が少なく、収集対象に偏りがある場合



5 データを分析する (Analysis)

フェーズ

Pr

Pl

D

A

C

Analysis

2. 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.53にて解説しています。

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

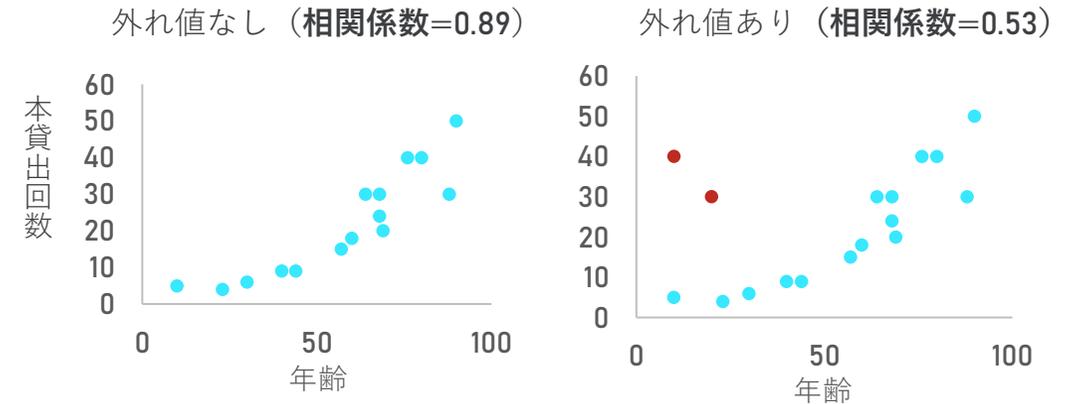
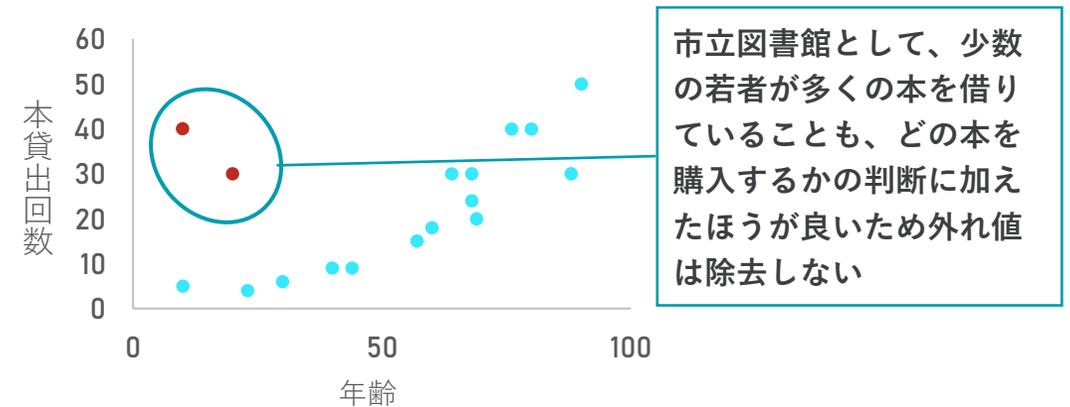


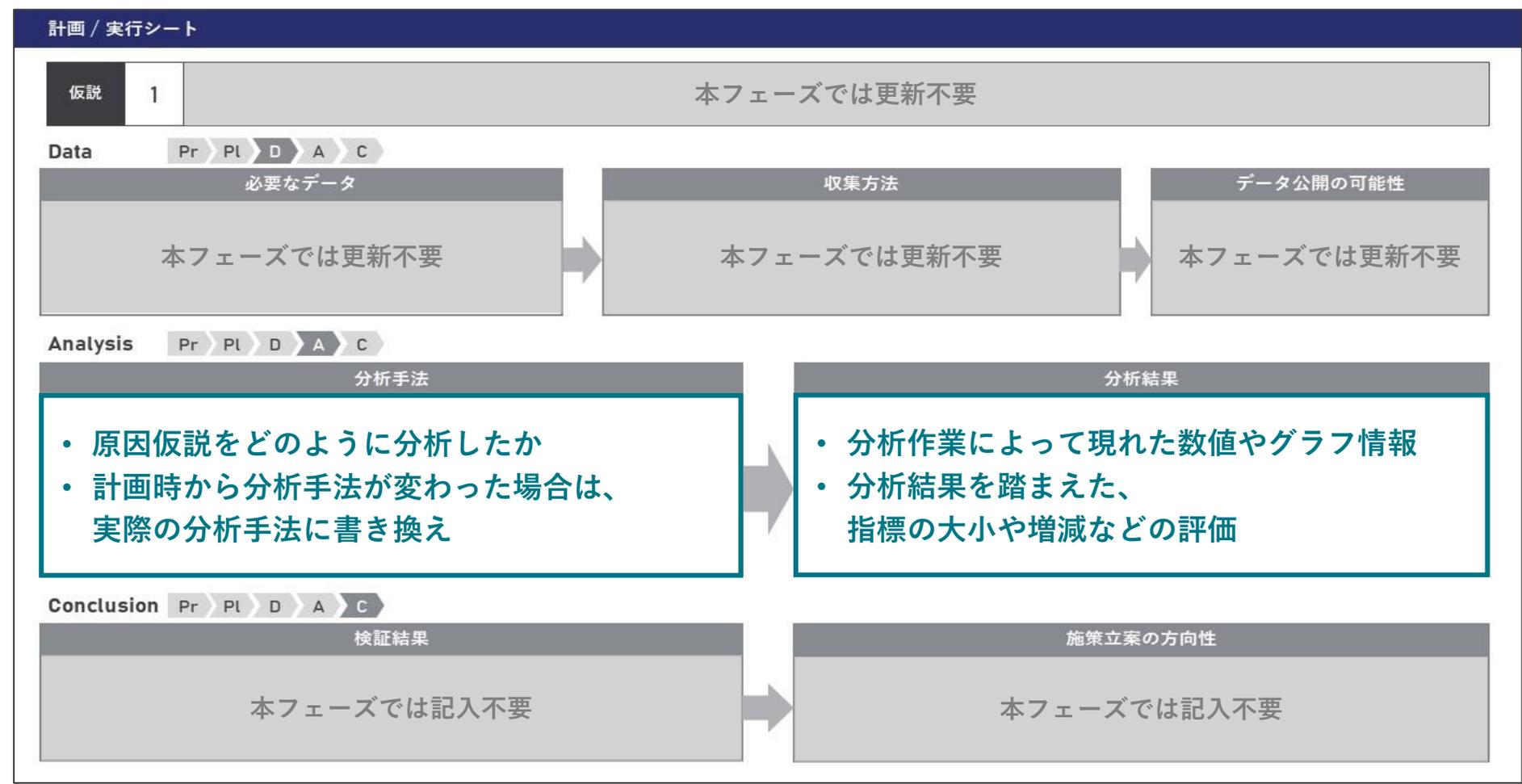
図2 外れ値を取り除かないほうが良い場合



5 データを分析する (Analysis)

3. ワークシートでの使用方法

Analysisフェーズでは実際に行った分析内容や得られた分析結果をワークシートに記入します。計画時の内容を更新しましょう。



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するフェーズ

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の妥当性を判断し、分析の目的に立ち返ってどのように結果を業務に活かすべきか、次に取るべきアクションは何か、検討していきます。

PPDACワークシート (P.54で詳細説明)

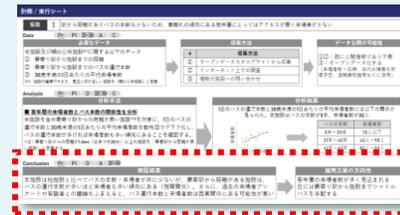
Conclusionフェーズでは検証結果を記入し、当初立てた仮説が正しかったかどうか判断します。また、検証結果を踏まえて施策の方向性を記入します

検証結果

分析結果から得られる示唆を記入し、仮説を検証できたか記入します

施策の方向性

原因を踏まえて次に提案する施策の方向性や、検証する仮説を記入します



必ず守るポイント

- 分析結果からどのようなことが言えそうか（示唆）を検討したうえで、当初考えていた原因仮説と分析結果を比較し、仮説が正しかったかどうか（妥当性）を判断しましょう
- 原因仮説が正しいと結論づけられた場合、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。想定と違う結果が出た場合でも、その結果をもとに次のアクションを検討しましょう

重要な考え方

- ✓ 仮説の妥当性を判断する場合は、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、周囲の人の意見を取り入れながら客観的に考えることが重要です
- ✓ 分析結果を考察する際には、相関関係と因果関係を混同しないよう注意することが大切です

P.53

1. 仮説が正しいか判断し、業務に活かす

Conclusionフェーズでは、仮説の妥当性を確認したうえで業務への活用を検討します

仮説が正しかったかどうかを判断する

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を比較し、原因仮説が正しかったかどうか（妥当性）を判断します。ただし、これまで実践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る可能性があるため、周囲の人の意見を聞きながら客観的に考え、判断することが重要です。

よくある例として、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



満足度調査結果が芳しくないが、一時的なものだろう



他職員と論理に矛盾がないか確認したり、他の示唆も得られないか議論しましょう



満足度が低い地区に絞って他のデータも見よう、別の問題が見つかるかも！

分析結果を業務に活かす

原因仮説が正しいと結論づけられた場合は、当初の背景や問題に立ち返って取組を検討する・裏付けとして整備するなど、分析結果を業務に活かしましょう。一方、仮説と異なる結果が出たとしても「原因はここではなく他のところにありそうだ」と分かったこととなります。データから得られた新しい情報をより良い施策や業務に繋げましょう。



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！



6 分析結果を考察する (Conclusion)

💡 COLUMN 相関関係と因果関係

Conclusionフェーズにて原因仮説が正しかったどうかを検証する際、原因仮説と起きている事象の相関関係に着目する場合があります。

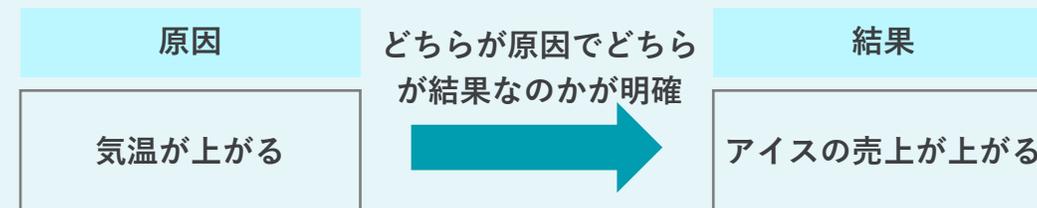
このとき、相関関係のある2つの変数には因果関係があるものもあります。因果関係とは、相関関係のなかでもAを原因としてBが変動する関係にあることを指します。

例えば「気温とアイスの売上」は、右図の①の通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。右図の②では、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があるか分かりません。相関関係と因果関係を混同しないようにしましょう。

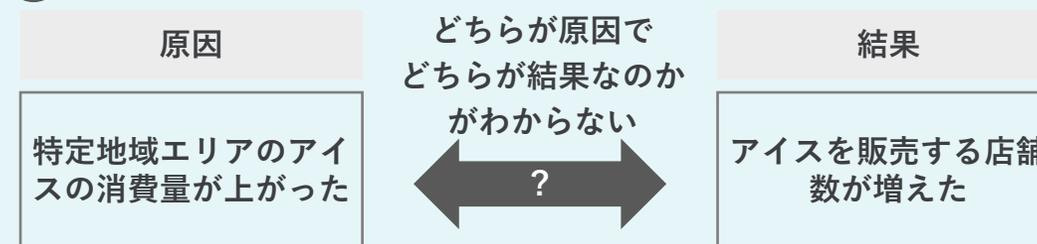
相関関係と因果関係



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

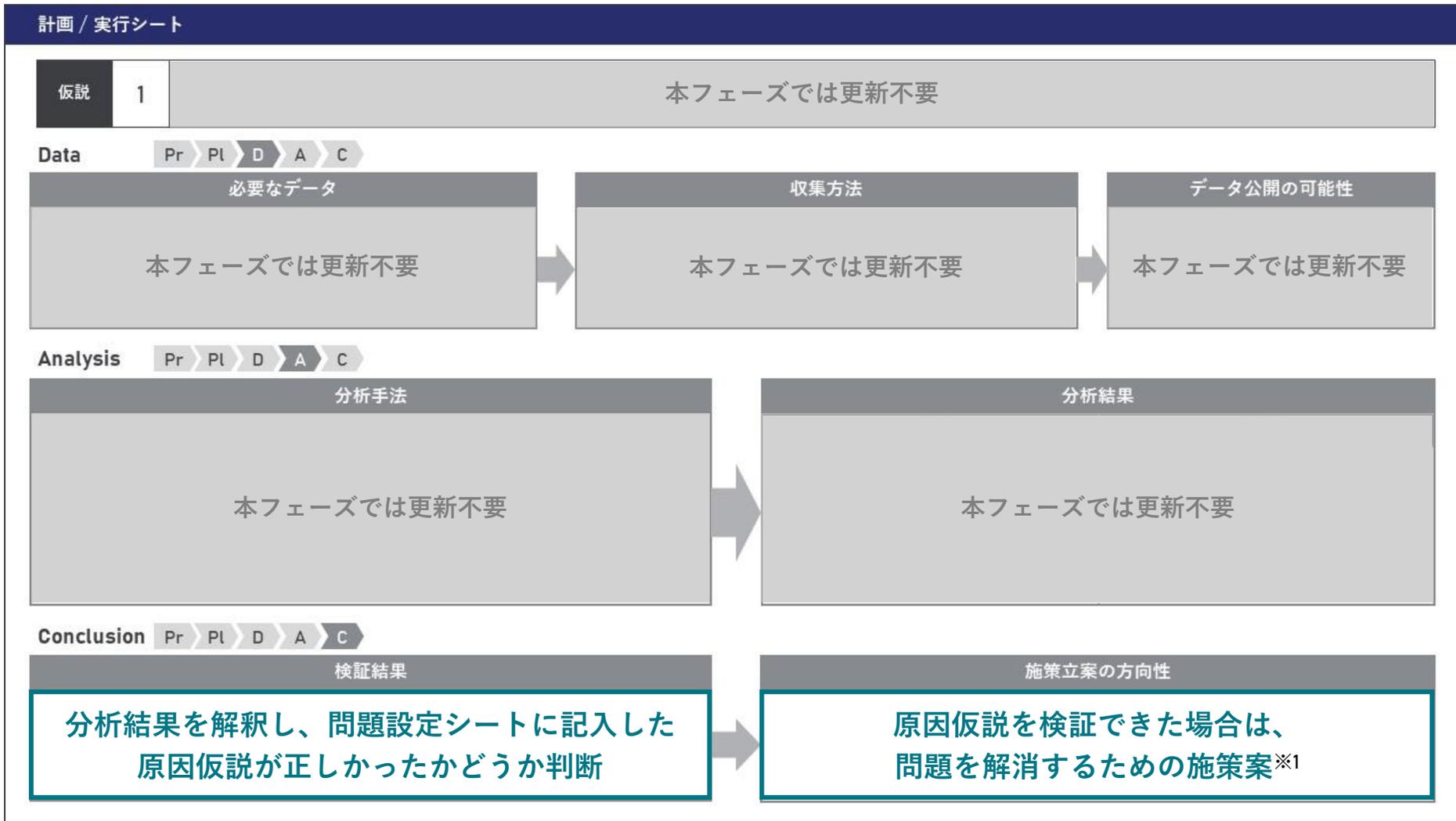




6 分析結果を考察する (Conclusion)

2. Conclusionフェーズにおけるワークシートのご使用方法

Conclusionフェーズでは、検証結果及び施策の方向性をワークシートに記入します。



※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します。

7 データ分析の実践例

1. ワークシートを活用した分析の実践例

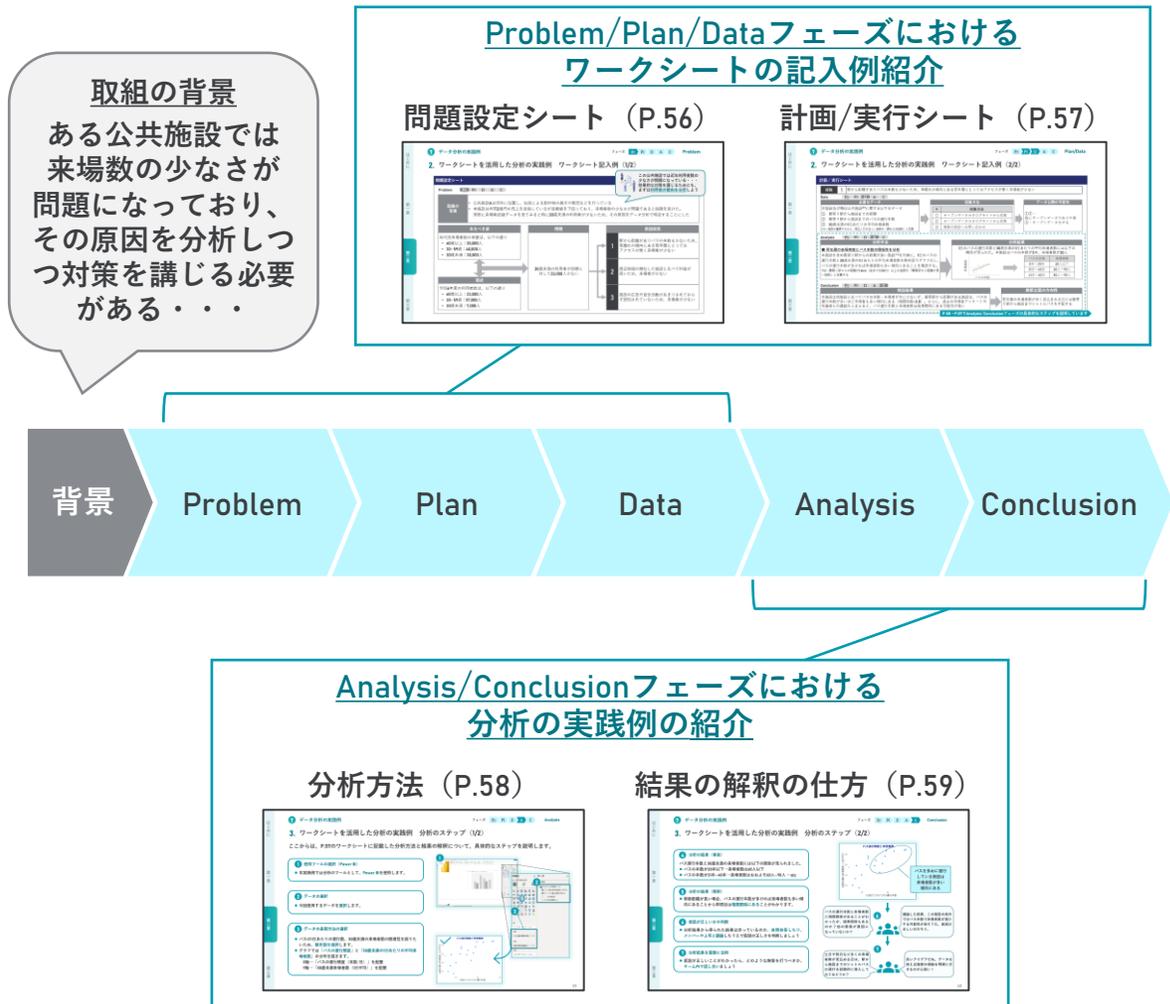
ここからは、PPDACワークシートを活用した実践例を説明します。

まず、P.56～P.57ではワークシートの考え方や記入例を紹介します。

次に、P.58～P.59ではワークシートの内容をもとにした分析の実践例を説明していきます。

日々の業務の中で、どのようにワークシートを活用し分析を行うのか、具体的にイメージしながら読み進めましょう。

分析の実践例の概要（P.56～P.59までの流れ）



7 データ分析の実践例

2. ワークシートを活用した分析の実践例 ワークシート記入例 (1/2)



この公共施設では近年利用者数の少なさが問題になっている・・・効果的な対策を講じるためにも、まずは**利用者の傾向を分析**しよう

問題設定シート

Problem **Pr** > **Pl** > **D** > **A** > **C**

取組の背景

- 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

あるべき姿

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6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

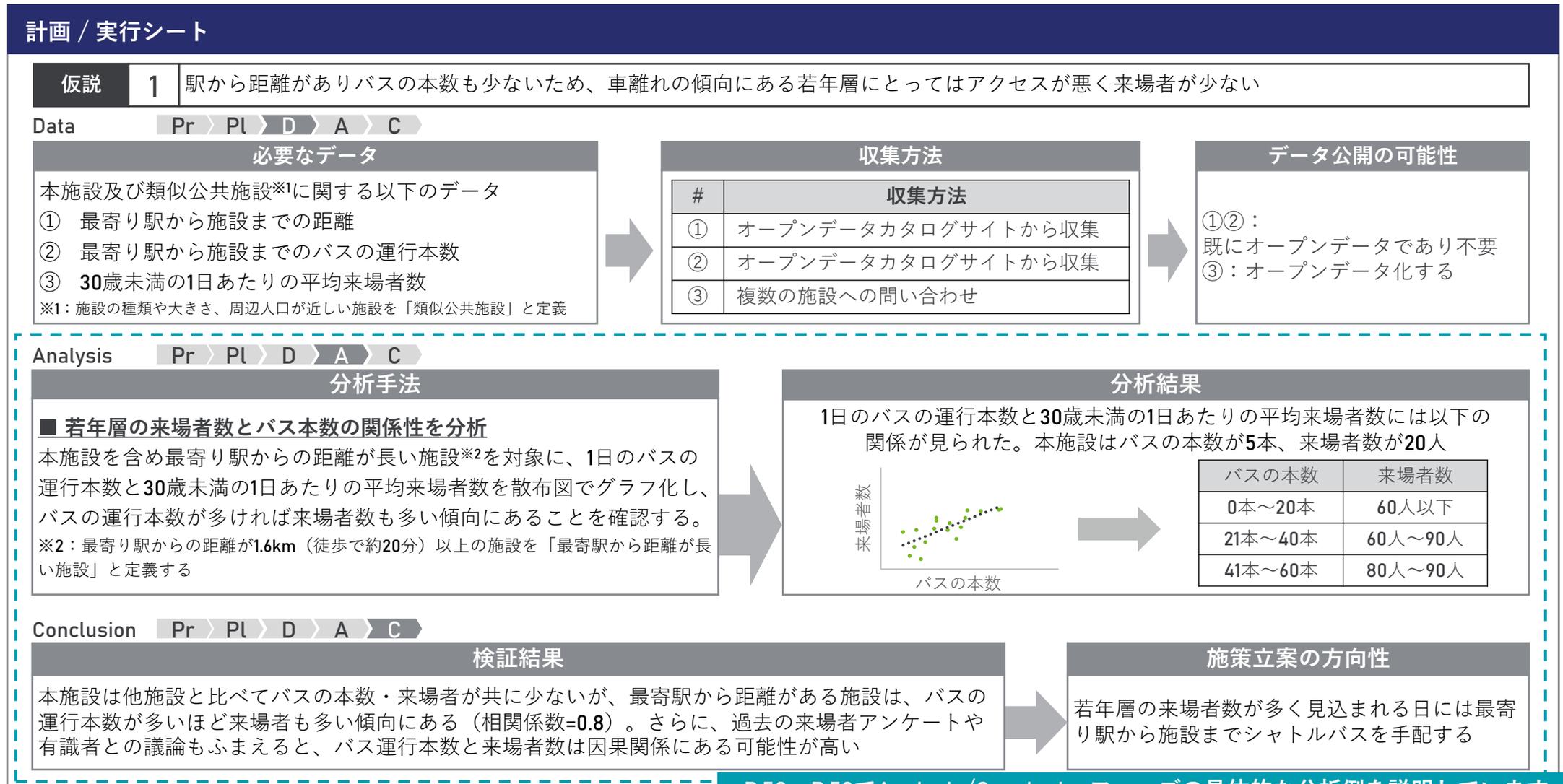
周辺地域の類似した施設と比べて料金が高いため、来場者が少ない

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

7 データ分析の実践例

2. ワークシートを活用した分析の実践例 ワークシート記入例 (2/2)



7 データ分析の実践例

3. ワークシートを活用した分析の実践例 (1/2)

ここからは、P.57のワークシートに記載した分析方法と結果の解釈について、具体的な分析例を説明します。

1 使用ツールの選択 (Power BI)

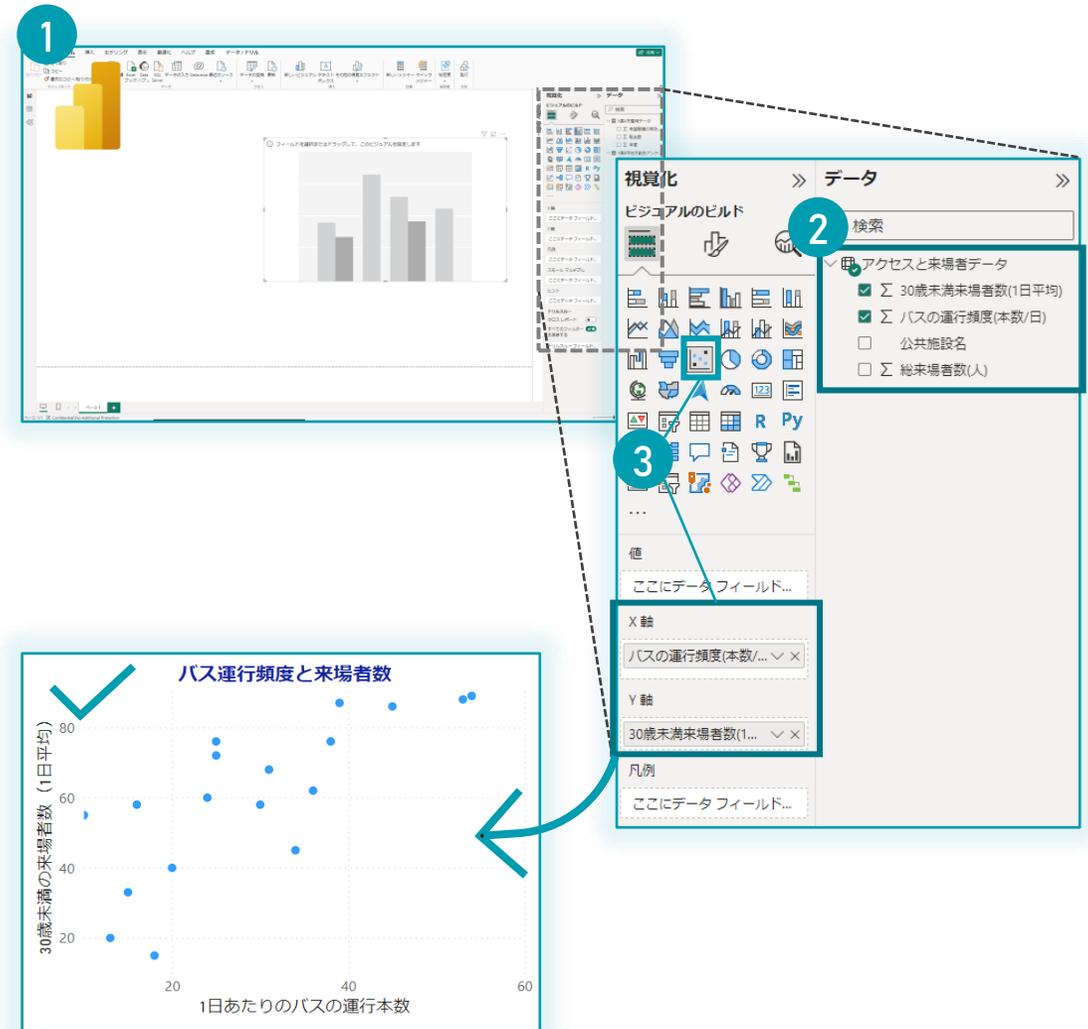
- 本実践例では分析のツールとして、Power BIを使用します。

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



7 データ分析の実践例

3. ワークシートを活用した分析の実践例 (2/2)

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60人以下
- バスの本数が21本~40本→来場者数はおよそ60人~90人 …etc

5 分析の結果 (解釈)

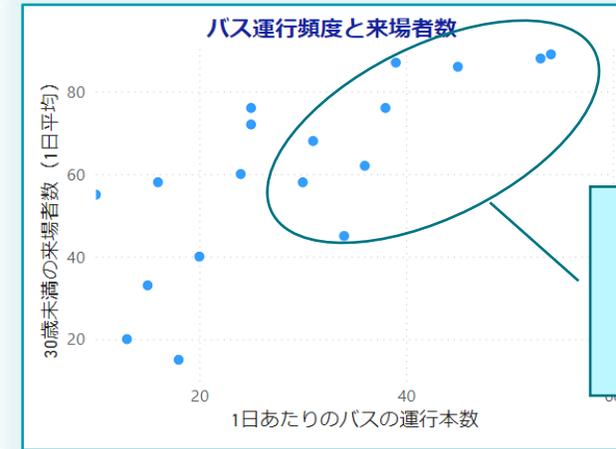
- 移動距離が長い場合、バスの運行本数が多ければ来場者数も多い傾向にあることから両項目は**相関関係にある**ことがわかります。

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

6
バスの運行本数と来場者数に相関関係があることがわかったが、因果関係もあるのか？他の要素が原因になっていないか？

議論した結果、この施設の条件ではバス本数で来場者数が減少する可能性が高そうだ。仮説は正しいのだろう。

7
土日や祝日など多くの来場者数が見込める日は、駅から施設までのシャトルバスの運行を試験的に導入して見てはどうか？

良いアイデアだね。データを使えば施策の根拠を明確に示せるのが心強い！

第三章： チェックリスト

第三章では、第二章 Dataフェーズ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.61 |
| ② | データ品質チェックリスト | ————— | P.62 |

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、「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。

📄 データ利活用チェックシート

Q 番号	データを提供するときの観点	実行を推奨する事項	チェック
Q1	提供データが、第三者に漏えいしてしまう場合に備えて、何か手立てはないか。	a. 契約で第三者提供の禁止やデータの管理方法の義務付け等を規定 b. システム設計による漏えい防止策を実施 c. データ漏えいの発生に備え、監視保全を実施	<input type="checkbox"/>
Q2	提供先が、提供データを自由に転用し、当社に不利な目的で転用する場合には、何か手立てはないか。	a. 契約で目的外使用の禁止等を規定 b. サンプルデータの提供やクラウドへのアクセス権限付与によるリスク軽減策を実施 c. 目的外使用が発生した場合に備え、監視保全を実施	<input type="checkbox"/>
Q3	提供先が、提供データを第三者に転用や権利化する場合には、何か手立てはないか。	a. 契約上、特許出願・権利化にはデータ提供者との事前協議を必要とする旨を規定 b. 特許法に基づき、特許移転請求等による対応	<input type="checkbox"/>

「データ利活用チェックシート」の概要及び主な項目

	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"> ✓ 契約で目的外使用の禁止などを規定する ✓ 個人情報の第三者提供の可否を確認する
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"> ✓ 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
データ使用	データを使用する 	<ul style="list-style-type: none"> ✓ 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 ✓ 契約における利用範囲を確認する
プラットフォーム	データを共有する 	<ul style="list-style-type: none"> ✓ 設計に際し、中立性・信頼性の確保を意識する

ワンポイント
 利用許諾取得の考え方については
 第二章DataフェーズP.32を参照ください

※1: 「データ利活用チェックシート」の詳細については、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⑤)を参照してください

2 データ品質チェックリスト

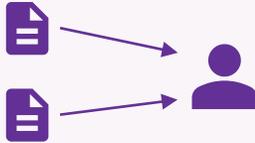
1. データ品質チェックリストの概要

データ品質管理を実践する際は、「データ品質チェックリスト」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。

次頁以降では、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」概要及び利用シーン

	利用シーン	チェック項目の概要
企画 P.63	PPDACに沿ってデータ分析を進める場合  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
要件定義 P.66	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
設計/開発 P.67	 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

2 データ品質チェックリスト

フェーズ

Pr

Pl

D

A

C

Data

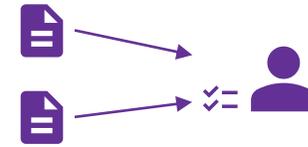
2. 使用方法（企画工程）

企画工程向けのチェックリストは、PPDACに沿ってデータ分析を行う場合、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、データ分析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へ委託する場面です。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へ委託する
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

 COLUMN Excelでデータを作るときにの心掛け (1/2)

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、データを分析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、後で利用しにくいいため控えましょう。





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。





項番	自治体	場所
XXX	芝浦小学校	
XXX	港南小学校	
XXX	港南中学校	
XXX	天神小学校	
XXX	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

数値に文字を混ぜると、Excel上では文字列として認識されます。





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。





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

COLUMN Excelでデータを作るときの心掛け (2/2)

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。

空白行でデータが分断

空白行を削除し一表

表を折り返し

折り返しせずに一表

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。

一部省略

省略せずに全て記入

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。

オブジェクトを使用している

オブジェクトを削除している

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。

2 データ品質チェックリスト

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、システム内及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

2 データ品質チェックリスト

フェーズ

Pr

Pl

D

A

C

Data

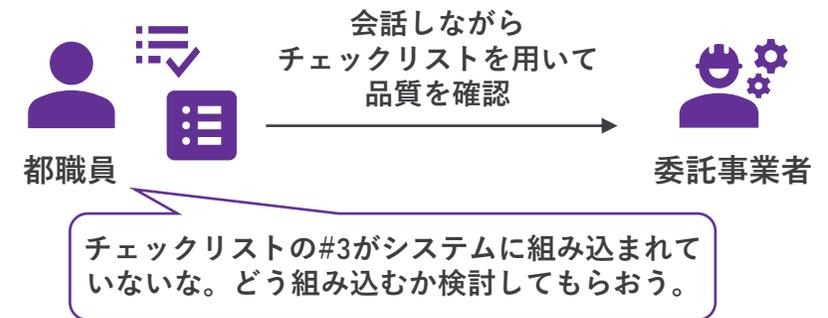
4. 使用方法（設計・開発工程）

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 セミナール編（1）～データ利活用の進め方」（<https://www.stat.go.jp/dstart/point/seminar1/>）2025年3月18日閲覧
- ② デジタル庁（2024）「ダッシュボードデザインの実践ガイドブック」（https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/1948e3cd-736a-4378-9e31-039b08d11106/10b4c14d/20240903_resources_dashboard-guidebook_guidebook_01.pdf）2025年3月26日閲覧
- ③ 文部科学省（2023）「「ロジックモデル」作成マニュアル」（https://www.mext.go.jp/content/20230410-mxt_kanseisk01-100000155-3.pdf）2025年3月26日閲覧
- ④ 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>）2025年3月18日閲覧
- ⑤ 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>）2025年3月18日閲覧
- ⑥ デジタル庁（2022）「政府相互運用性フレームワーク(GIF) 460 実践ガイドブック」（https://github.com/JDA-DM/GIF/tree/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）2025年3月18日閲覧
- ⑦ デジタル庁（2024）「政府相互運用性フレームワーク（GIF）」（https://www.digital.go.jp/policies/data_strategy_government_interoperability_framework）2025年3月18日閲覧
- ⑧ デジタル庁（2023）「自治体標準オープンデータセット（正式版）」（https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test）2025年3月18日閲覧
- ⑨ 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）2025年3月18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
P.5	東京都デジタルサービスの開発・運用に係る行動指針	https://www.digitalservice.metro.tokyo.lg.jp/documents/d/digitalservice/02_koudoushishin
P.12	東京都オープンデータ推進庁内ガイドライン	https://portal.data.metro.tokyo.lg.jp/guideline/
	データ整備事業データ入力マニュアル 庁内データ連携基盤のデータ活用ガイドライン	https://www.digitalservice.metro.tokyo.lg.jp/documents/d/digitalservice/dataentry_manual (都庁職員のみ閲覧可能)
P.14	都税収入見える化ダッシュボード	https://app.powerbi.com/view?r=eyJrljoiZjUzNmVjMTAtMjA0Ni00ODI0LWI1ODctMDI2MWEzNDIyMThkliwidCl6lmQwMzAyZmNjLTNlODEtNDIjMy04MjM1LWQzMTEhMzY4NGNmYyJ9&pageName=ReportSection862d2d37bcc6ca2a0120
	東京地中熱ポテンシャルマップ	https://www.tokyogeoheatmap.metro.tokyo.lg.jp/
P.30	東京都デジタルツイン実現プロジェクト	https://info.tokyo-digitaltwin.metro.tokyo.lg.jp/
P.38	東京オープンデータカタログサイト	https://portal.data.metro.tokyo.lg.jp/
	東京データプラットフォーム	https://www.tdpf.metro.tokyo.lg.jp/

その他参考文献

- ・内田康雄（2021）「Tips for creating machine-friendly data」 (<https://www.slideshare.net/uchiruda/tips-for-creating-machine-friendly-data-248311530>) 2025年3月18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) 2025年3月18日閲覧
- ・総務省「e-Stat 政府統計の総合窓口」 (<https://www.e-stat.go.jp/>) 2025年3月18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) 2025年3月18日閲覧
- ・総務省（2019）「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) 2025年3月18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) 2025年3月18日閲覧
- ・総務省（2024）「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) 2025年3月18日閲覧
- ・総務省（2019）「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) 2025年3月18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) 2025年3月18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>) 2025年3月18日閲覧
- ・デジタル庁「介護現場の生産性向上に関するダッシュボード」 (<https://www.digital.go.jp/resources/govdashboard/nursing-care-productivity>) 2025年3月18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) 2025年3月18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>) 2025年3月18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) 2025年3月18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