

総務省統計局統計データ利活用センター主催

EBPMブートキャンプ

〈データ分析ワークブック〉



研究会メモ

EBPMブートキャンプ「ワークブック」

ワークブック前半では、まず課題に関する現状把握をしっかりと行います。その後、以下①～⑤の順で思考し、最終的に分析に必要なデータの洗い出しを行います。

例1

②KPI / KGI

KPI:年代別糖尿病症患者数
KGI:〇年までに糖尿病罹患患者数を〇%減少

①分析目的

〇〇市特有の糖尿病予備軍の傾向を把握し、市による注意喚起の対象となる層を見つけ出す

③仮説

長時間労働が関係してるのではないか

④分析のアプローチ

-集計項目・必要なグラフ

職業別の罹患率と職業別平均週間労働時間の比較グラフ等

⑤データ

-分析に必要なデータ

職業別の糖尿病罹患率、職業別労働時間データ等

例2

②KPI / KGI

KPI: 25歳～44歳の女性就業率
KGI:〇年までに25歳～44歳の女性就業率を〇%以上にする

①分析目的

〇〇市の25歳～44歳の女性の就業率を上昇させたい

③仮説

子育て(介護)との両立が難しく、退職してしまう人が多いのではないか

④分析のアプローチ

-集計項目・必要なグラフ

25歳～44歳女性の退職理由の内訳がわかる帯グラフ等

⑤データ

-分析に必要なデータ

25歳～44歳女性の正社員退職理由についてのアンケートデータ等

現状把握

データ利活用プロジェクトでは、研究者全員が共通認識をもつことが最重要です。まずは、課題にまつわる状況や抱えている問題・課題について、思いつく限り挙げてみましょう。

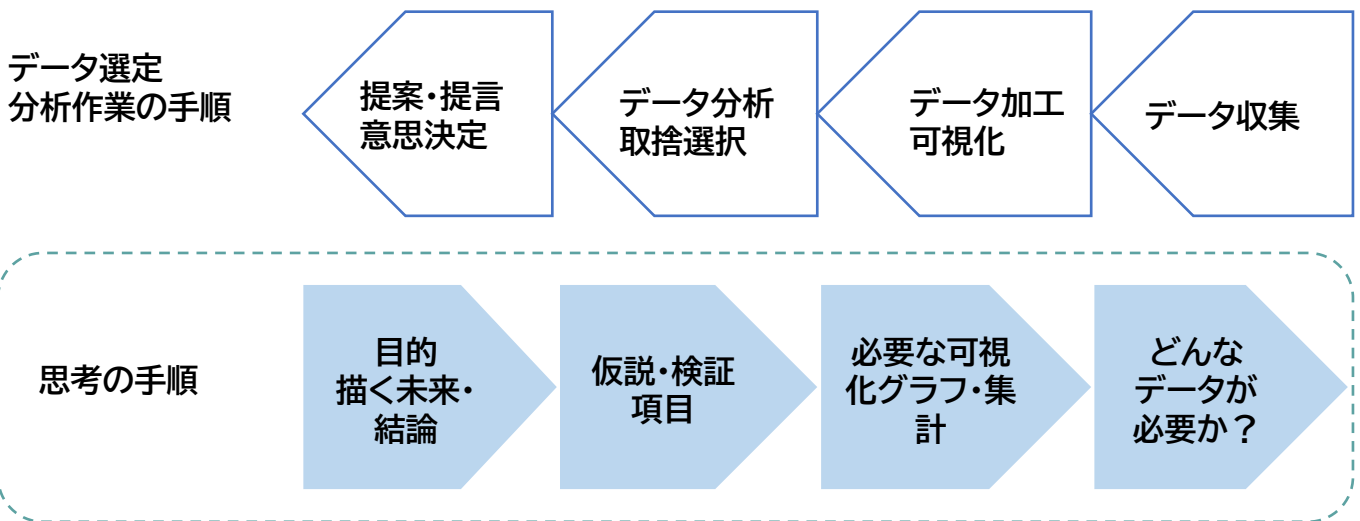
<現状把握の記入例>

- ・市内で働いている女性のうち、出産後に退職してしまう人が20%ほどいる

データ利活用のプロセスとは

データ利活用プロジェクトにおいては、「仮説」「ストーリー」を組み立ててから、それらを裏付けるデータを集めます。そうでなければ「意味のないデータ」を処理することになってしまいます。はじめに高い解像度で目的を定めることで、より有効にデータを活用することができます。

【データ利活用のプロセス】

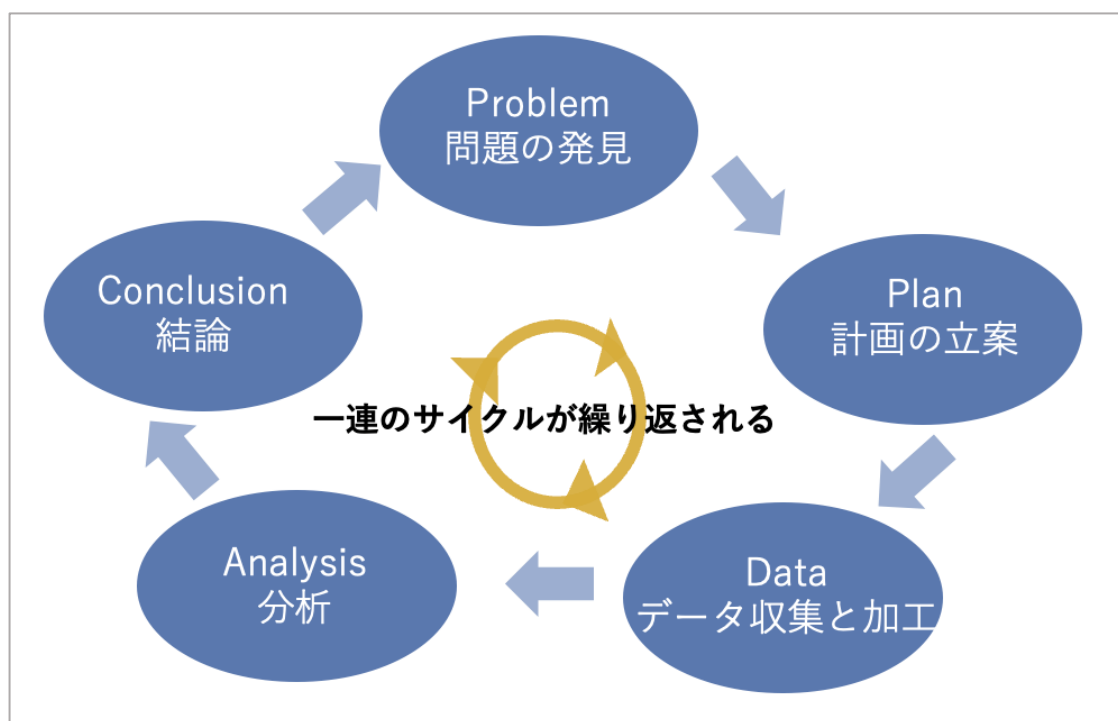


データありきではなく、目的ありきで考えることが重要です。まずは、目的を明確にすることから始めましょう。



データサイエンスのサイクル(PPDAC)とは

PPDACサイクルとは問題を解決するためのプロセスの一つです。問題解決における各段階として、Problem(問題)、Plan(計画)、Data(データ)、Analysis(分析)、Conclusion(結論)に分割します。



課題解決においては、PPDACサイクルの一連を繰り返し行うことが大切です。

多くの場合、一度の仮説検証の結果から、新たな課題が見出されます。さらなる深堀りのための分析計画を立案し、PPDACサイクルを繰り返すことで、当初の課題の解決策がより鮮明になります。



③仮説

仮説立案とは

課題を整理したのちに、仮説立案を行います。仮説とは、ある現象を説明するために確からしいといえる仮の結論です。これを基にデータ収集やデータ解析を行い、仮説の検証をしていきます。

仮説立案を行うことで、全体でどのくらい仮説があるか、また注力すべき仮説はあるかが明確になります。

したがって、非常に重要なステップといえます。

<仮説立案のヒント>

・具体的に

⇒なるべく具体的に落とし込む

・網羅的に

⇒ MECE(もれなく、重複なく)の視点を忘れずに

例)ロジックツリー等の活用

⇒ 仮説と反対の観点から考え直してみる

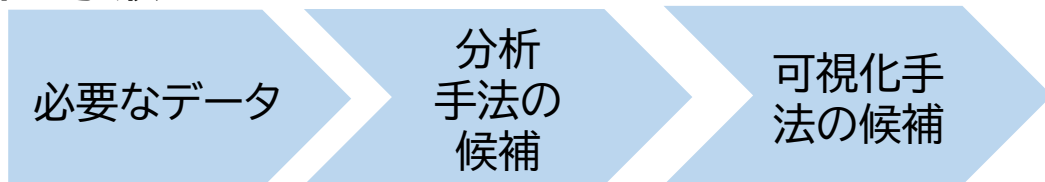
研究課題:

④分析のアプローチ(集計項目・必要なグラフ)

仮説検証のアプローチ設計

仮説立案を行った後、分析のアプローチ設計を行います。分析アプローチ設計は、思いつきや手当たり次第の分析、作業の出戻りや無駄な分析を行なわないために必要な作業です。

<設計の手順>



分析に必要な集計表とは??

データ分析の集計を行う際には、分析目的にあった集計の切り口、比較軸で集計を行うことが重要です。「売上の向上」が目的の場合、顧客の属性購入された商品の種類をクロス集計することで、顧客の傾向を把握することができ、プロモーションを考える際に役に立つ集計表となります。分析に必要な集計表のイメージを書いてみましょう。

データの可視化

データの可視化の目的は、数値情報だけでは捉えにくい現象や傾向などを把握しやすくするためです。例えば、売上と降水量のデータをグラフにすることで、降水量が少ないときには売り上げが多いというような2変数の関係性を捉えることができ、データからの事象を解釈するための手助けになります。

⑤データ(分析に必要なデータ)

分析に必要なデータ

課題や仮説を設定したのちに、その課題や仮説の原因や解決策を探るために必要なデータを選定します。例えば、売上の向上を目的とすると、日別、商品別の売上、客単価、商品単価などのデータが分析が必要になると考えられます。このように、分析の目的にあったデータを選定することが重要です。まずは、分析に必要なデータとデータ項目を挙げてみましょう。

必要データ	項目	期間

必要なデータの期間

時系列でデータの分析を行なう場合、集計データの期間が十分でなければノイズの影響を受けやすく、また長すぎると需要特性が変化したり、変動要因が分かりにくくなります。季節変動、トレンド、ノイズなどを考慮し、データの種類、分析の目的に適したデータ期間を設定することが重要になります。先ほど挙げた必要データにおいて、必要なデータの期間を検討し、上の表の右欄に記入しましょう。

データの取得方法

分析に必要なデータを収集する方法として、e-statなどのweb上で提供されているデータや、ウェブAPI、社内のデータベースを利用する方法があります。また、データが存在しない場合には、新規で調査をするなどし、データを取得することもあります。

必要データで挙げたデータの取得方法と取得にかかる費用について、以下の表にまとめてみましょう。

必要データ	取得方法	費用

活用範囲の制約事項の確認

取得したデータによっては、データ利用の制約事項があります。特に、個人情報を含んだデータを利用する際には、個人情報保護の観点から、一定の匿名加工、取得時の本人同意などが必要になります。

データの制約事項について問題がないか確認しましょう。

データの収集

収集が必要なデータに対して、実際に収集できたもの・できなかったものを記入しましょう。
収集できたデータに時系列データがある場合は、期間を記入しましょう。

収集できたデータ・期間	収集できなかったデータ

可視化目的の定義

数値データを可視化することで、データの全体像の解釈に役立ちます。
具体的には、次のようなメリットがあります。

- ・比較ができる
- ・データの規則性を見出しやすくなる
- ・情報共有がしやすくなる

可視化のイメージ

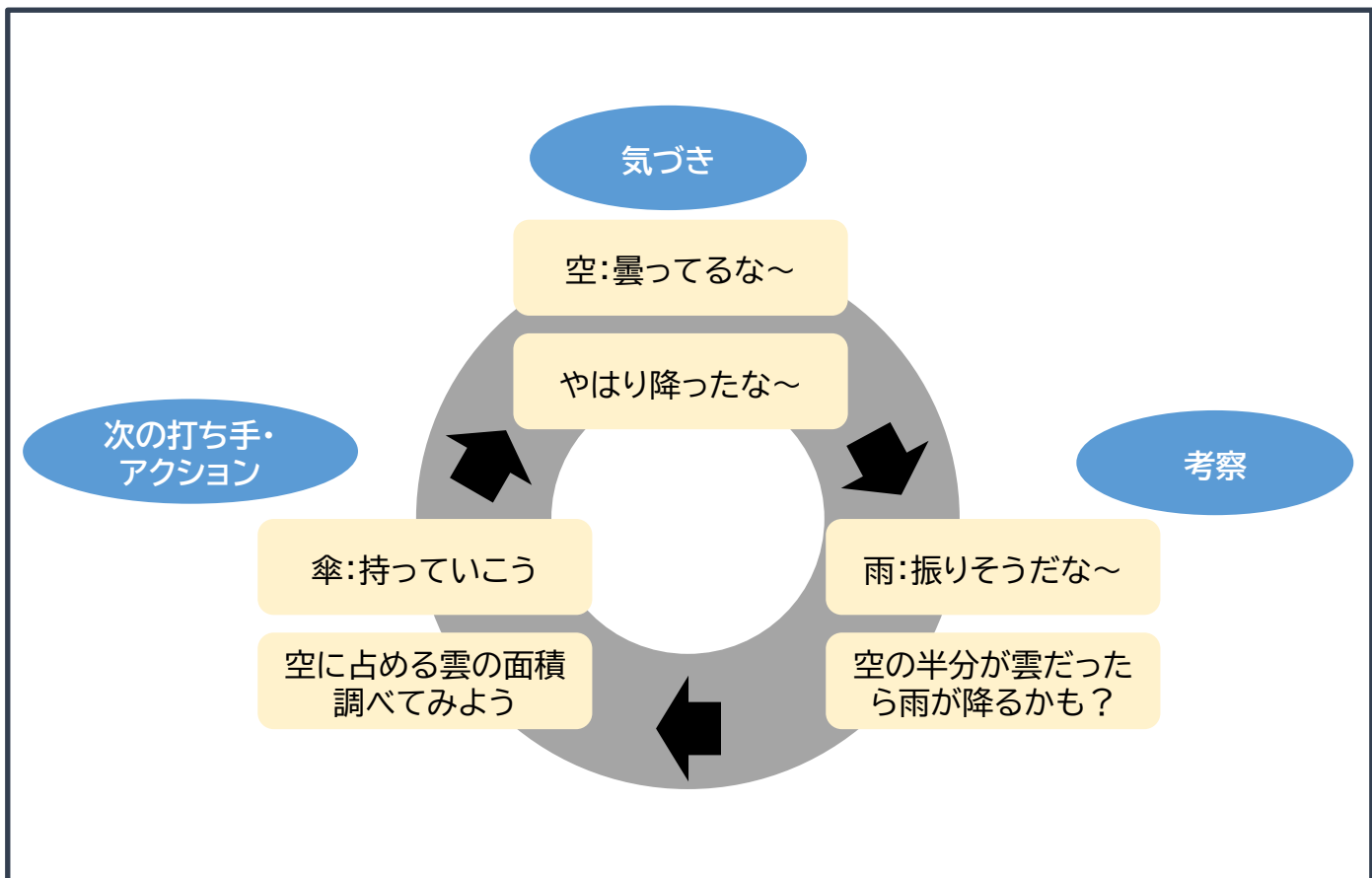
収集したデータをどのような手法で可視化するか記入しましょう。

<記入のヒント>

可視化方法によって、どのような気づきを得られそうか考えてみましょう。

例)曜日ごとの売上個数の棒グラフ、月ごとの平均気温の折れ線グラフ

空・雨・傘のサイクル



データを見た上での気づき

データから意味合いを見抜く方法として、「いかに気づくか」と仮説検証をする「空・雨・傘」のサイクルを回すことが有効です。

分析するデータから気づきを得るテクニックとして、次のことが挙げられます。

・特徴的なポイントを見つける

折れ線グラフのスパイク/データが無いところ/クロス集計表は大きい・小さいところ

・パターンを見つける

時系列グラフは「周期性(トレンド)」、散布図の方向性

・見えてくる線を引く

仮説や検証アイデアが生まれやすい

・想定を書き足す

こうだったはずでは?

こうなったらいいのに? 解決策にもつながりやすい

空・雨・傘の実践

「空・雨・傘」サイクルを使って、データを読み解いてみましょう。

空(気づき)	雨(考察・仮説)	傘(次の打ち手)

行き詰ったら、データを見る際の
テクニックを思い出しましょう。

- ・特徴的なポイントを見つける
- ・パターンを見つける
- ・見えてくる線を引く
- ・想定を書き足す



PPDACサイクルの実行方法

問題解決のためのフレームワークであるPPDACサイクルでは、分析結果をもとに仮説的に結論を出し、分析、結論から得られる新たな課題について深掘りのため、再度サイクルを回していくというように繰り返し実行していくことが重要です。仮説検証を繰り返しおこなうことで、本来の課題の解決策に近づいていくことができます。

仮説に対する考察

PPDACの結論フェーズでは、分析結果から立案を行った仮説に対して、「考察」を行います。考察では、分析結果の解釈、立案した仮説が正しかったのかどうかに加え、分析から得られた新たな問題に対し、次のPPDACサイクルの立案や今後の分析方針なども考えていきます。実際の分析結果を見ながら考察を行ってみましょう。

<結論フェーズの手順>

1. 分析結果の解釈
2. 仮説に対しての判断
3. 問題の解決案を提案、新たな問題に対しての分析計画立案

Next Step

「データサイエンスのサイクル(PPDAC)とは」でも触れたとおり、課題解決においてはPPDACサイクルを繰り返すことが大切です。

今回の研究課題を掘り下げるためには、どのような追加の分析が考えられるか、ここまでの仮説検証を振り返りながら書き出してみましょう。

追加必要データ

再分析を実施するにあたって必要なデータを書き出してみましょう。

必要データ	項目	期間

データ分析レポートの書き方

データ分析レポートとは、データ分析を行った結果得られた知見をテキストや図表を用いて文書にまとめたものです。

一般的な構成

課題背景

分析対象テーマの課題やその背景、分析を進行する上での留意点を提示

アプローチ

使用したデータや、データの入手方法など、実行したアプローチを簡潔に記載

検討結果

分析結果と考察、仮説検証の結果を提示

意味合い

分析結果から得られた知見やさらなる仮説の記載

ネクストステップ

追加検討事項や解決策の提示

可視化のテクニック[おまけ]

<グラフ選択>

複数のグラフを比較する場合、グラフの種類を揃えましょう。

<必要事項の記入>

グラフに加え、表題(グラフテーマ)、縦軸、横軸、単位、出典名等を記入しましょう。

