

Contents

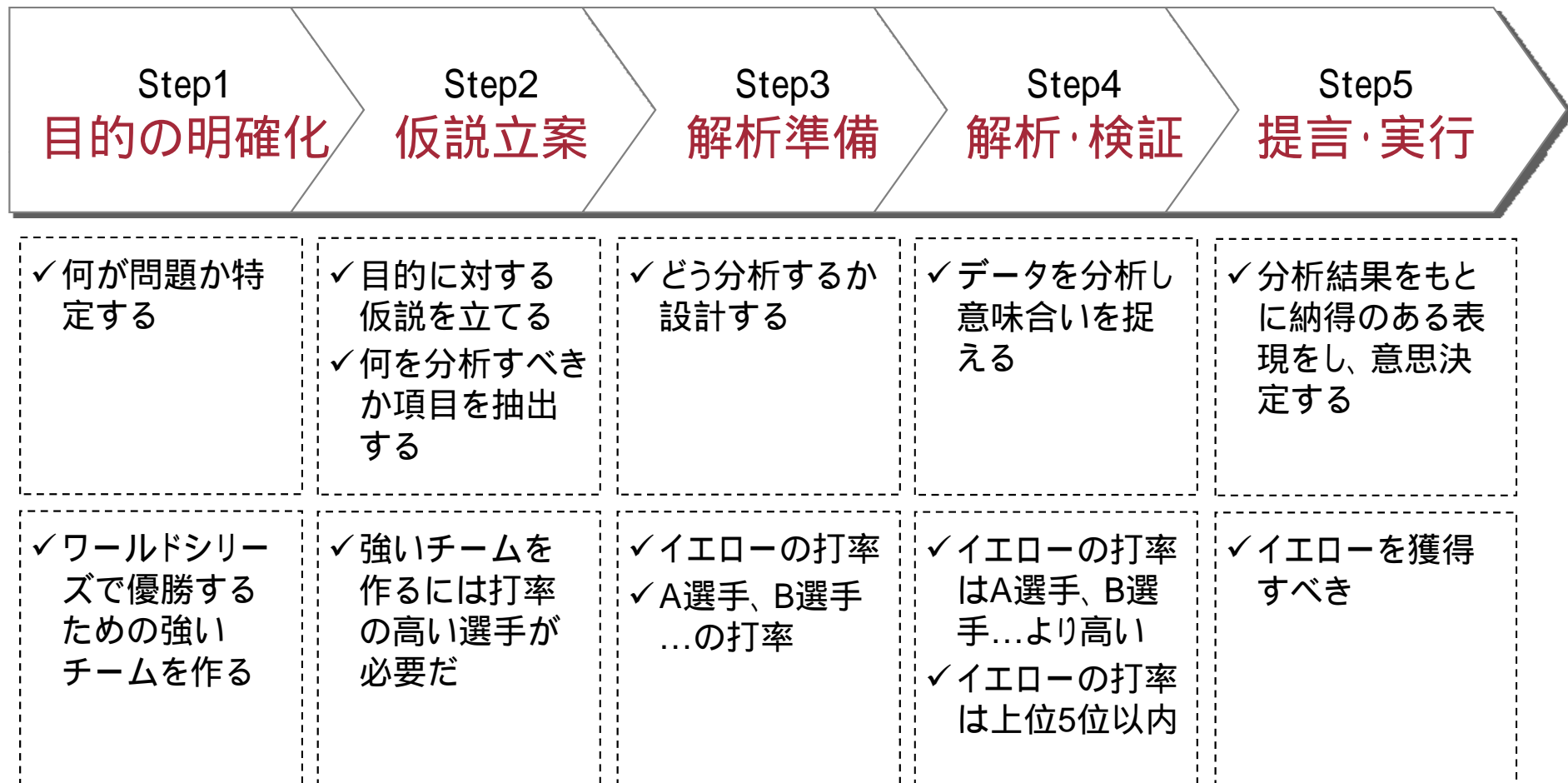
1. はじめに
2. 目的の明確化と意味合いの抽出の重要性
3. 仮説立案の重要性と合理的決定の方法
4. 仮説立案、解析準備、解析・検証のサイクル
5. そして実務へ



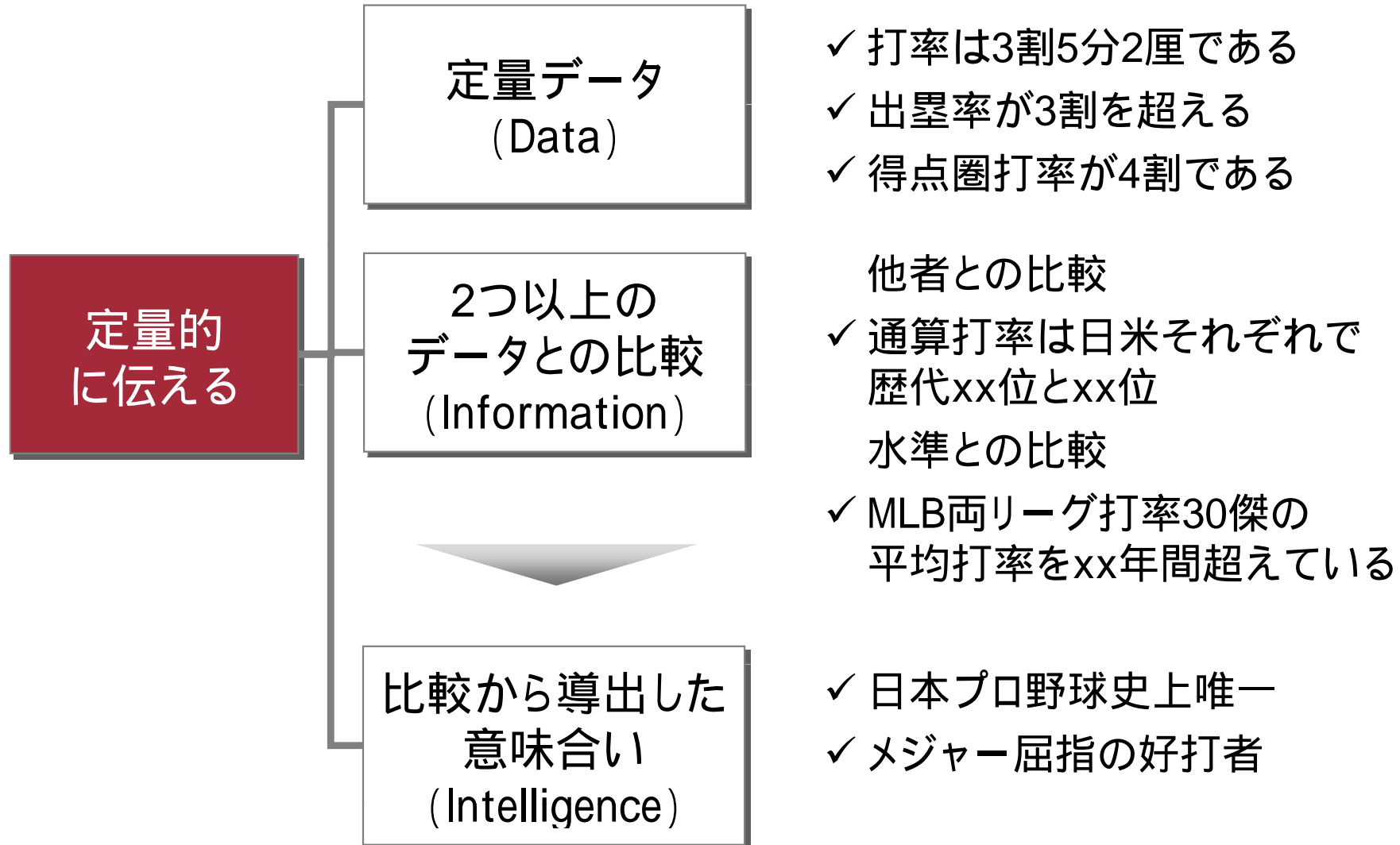
1-8. 定量分析のプロセス

定量分析には目的を明確化し精度・合理性の高い解決策を立案と数値を持って周囲を説得するための一連のプロセスがあります。

定量分析のプロセス

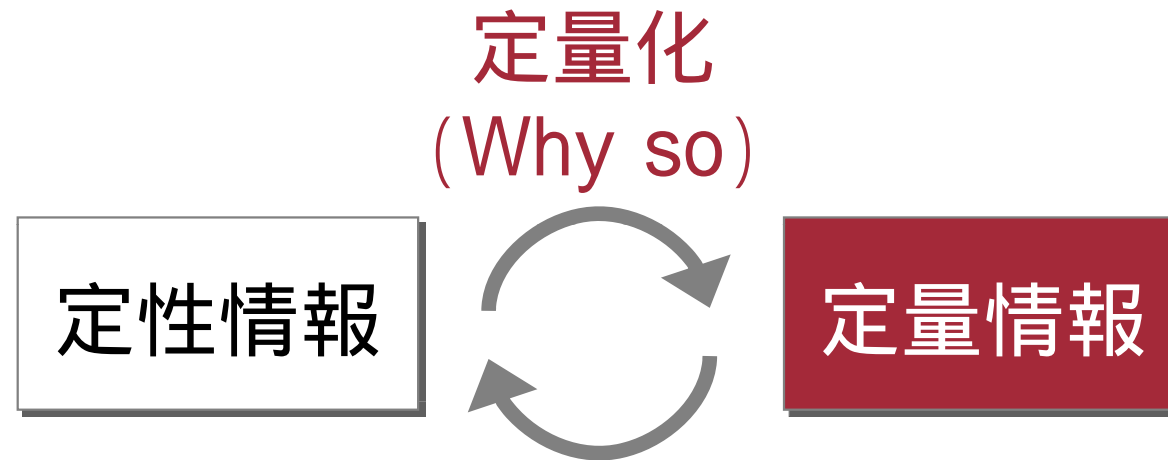


定量的に伝えるためには、単に定量データを示すのではなく、他者の数値や、水準、指標と比較することで、単独の数値が持つ意味合いを捉えることが重要です。



2-5. 定量化と意味合いの抽出の関係

定性情報と定量情報を活用するには、「定量化」と「意味合いの抽出」が重要であり、両者には密接な関係があります。そして、この関係を明確化することが重要です。



- ✓ 日本プロ野球史上唯一
- ✓ メジャー屈指の好打者

意味合いの抽出
(So What)

- ✓ 打率は3割5分2厘である
- ✓ 出塁率が3割を超える
- ✓ 得点圏打率が4割である
- ✓ 通算打率はMLB歴代xx位
- ✓ 両リーグ打率30傑の平均打率を10年間超えている

2-14. 目的の明確化の重要性

「何のために定量分析をするのか」を捉えること(目的の明確化)が、定量分析の初期段階では非常に重要です。目的が明確化されてはじめて、意思決定のために求められる要件を捉えること(仮説立案)ができます。

誤った定量分析

- ✓ とりあえず手元にあるデータをいじってみる
- ✓ 各データは全て同列で関係性や重要度は考慮しない
- ✓ データを加工して、それらしくまとめて終わり

スタンスが明確にみえない
安易な答えを
分析の結果とする

正しい定量分析

- ✓ 分析の目的を明確にする
- ✓ 目的を構造的・網羅的に分解し、全体像を捉える
- ✓ 目的に沿った仮説を立て、数値から検証・判断する

目的を押さえ続けることで
必要な分析の
イメージを明確にする

2-15. 意味合いの抽出の重要性

定量的な情報は意思決定のために求められる要件と結び付いたメッセージとなっ
はじめて意味合いをもちます。

誤った定量分析

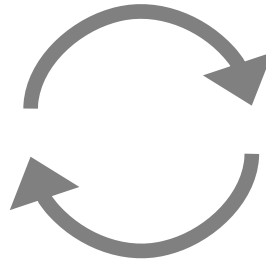
- ✓ たまたま手元にある情報で無理やり意味付ける

正しい定量分析

- ✓ 数値比較を通じて目的に向かったメッセージを抽出する

定量化
(Why so)

定性情報



定量情報

意味合いの抽出
(So What)

時系列に並べることで、その傾向から変動要因を検討します。一般的に「傾向変動」「循環変動」「季節変動」「不規則変動」に分解されます。また、傾向から将来を予測するにも非常に有効なものです。

概要

利用シーン

- ✓ 要素を時間の軸(月、四半期、年...)で捉え要素の変化を見たいとき

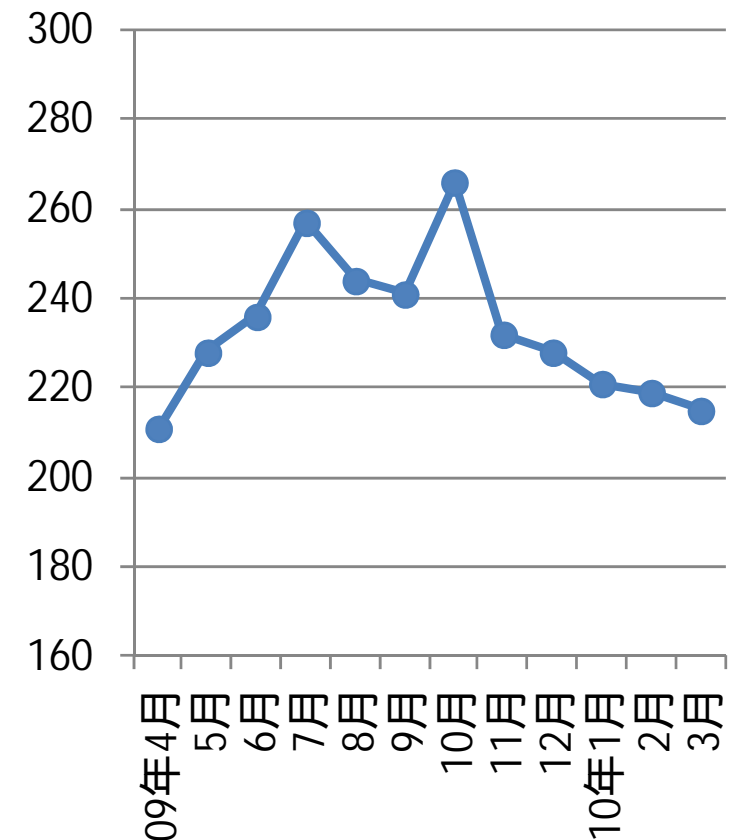
メリット

- ✓ 季節変動、循環変動などを除いた基底に流れる全体的な(本質的な)傾向が見える

注意点

- ✓ 見えた傾向が定性的な情報から推測される姿と整合しているかを確認すること

グッズ売上
(百万円)



全体を代表する数値として「平均値」を用いることは非常に頻繁に行われますが、使い方(特に母集団)に納得感があるか注意が必要です。

概要

利用シーン

- ✓ 全体の傾向を大まかに掴みたいとき

メリット

- ✓ 集団の代表値として直観的に分かりやすい

注意点

- ✓ 母集団の偏り、サンプル数が妥当か確認が必要

単純平均

- ✓ 何個かの数を全部加え個数で割ったもの
- ✓ 日経平均株価

加重平均

- ✓ 値の重みを加味して平均する
- ✓ 東証株価指数
時価総額を発行株式数でウェイト付け

「平均値」は正規分布を前提とする。ほかには中央値や最頻値があります。

概要

利用 シーン

- ✓ 正規分布や平均値だけでは表現できないとき

メリット

- ✓ 平均では分からない実態を捉えることができる

注意点

- ✓ 母集団に含まれる異常値の取り扱いを決める必要がある

中央値

- ✓ 要素を並べたとき中央に位置する値
- ✓ 世帯貯蓄額

最頻値

- ✓ 最も頻繁に出現する値
- ✓ 最多販売価格帯

ヒストグラムとは、全体をある切り口の層(階級)に分けて、その層の度数を棒グラフにして、y軸の多少を比較させることを目的で「ばらつきの全体像を捉える」ことが可能です。

概要

利用
シーン

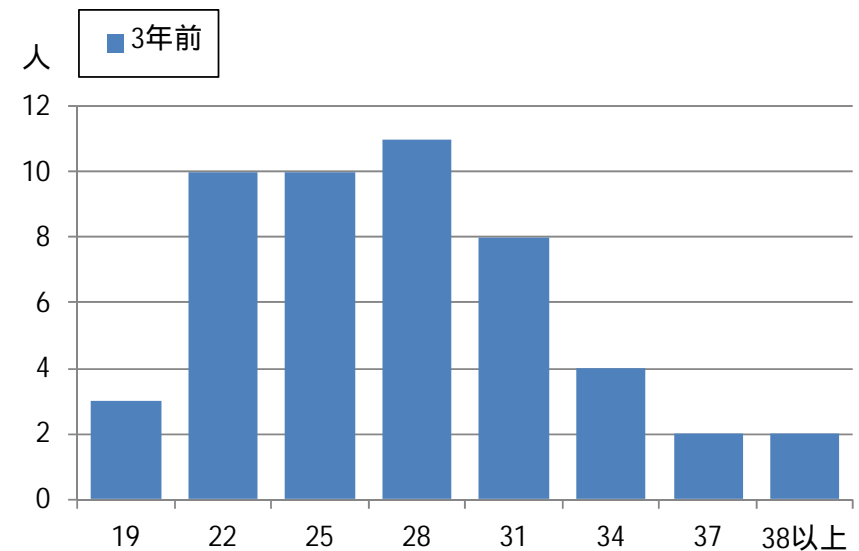
- ✓ 独立した多数の標本からなる集合に対して分布の全体感を見る

メリット

- ✓ ビジュアルな分かりやすさ
- ✓ 標本が集中しているところ、していないところが目瞭然

注意点

- ✓ 階級の取り方(数)次第で見映えが変わってしまう



パレート図は要素の大きい順に左から並べた棒グラフとその要素の左からの累計値が全体に占める割合(構成比率)の折れ線グラフと組み合わせたものです。

概要

利用
シーン

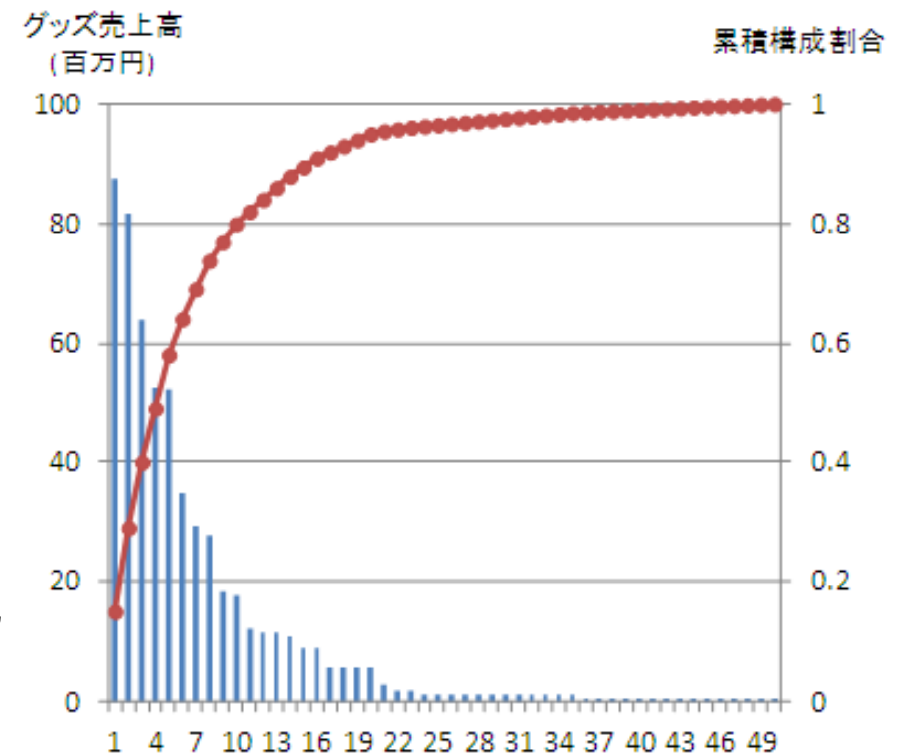
✓ 要素の順序を見たいとき

メリット

- ✓ 全体への影響が大きい要素が分かる
- ✓ 比較の視点がうまれる

注意点

- ✓ 要素がプロセスや時系列であったり、要素の独立性が低いときは適さない
- ✓ データ数が少ない場合はパレート図をつくる必要がない



クロス集計は2つの属性のデータを縦軸、横軸として集計したもので、二属性間に従属の関係があるかを見るものです。また、 χ^2 検定によって独立性の検定や決定木分析に活用されます。

概要

利用
シーン

✓ 2属性間の従属性/独立性を調べる時

メリット

✓ 集計するだけで、属性間の偏りを数字で確認できる

注意点

✓ 偏りが判断できるだけのサンプル数が必要となる

データの個数 / 生え抜き or 他チーム	列ラベル	行ラベル	生え抜き	他チームからの獲得	総計
A			2	8	10
B			5	5	10
C			19	11	30
総計			26	24	50

散布図は二変数をx軸、y軸としてデータをプロットしたもので、二変数に特定の関係(相関)があるか、いくつかの群に分けることができるかを見ることができます。また、散布図から回帰分析を行うことで、XとYの「相関の強さ」や「回帰式の精度」を評価できます。

概要

利用
シーン

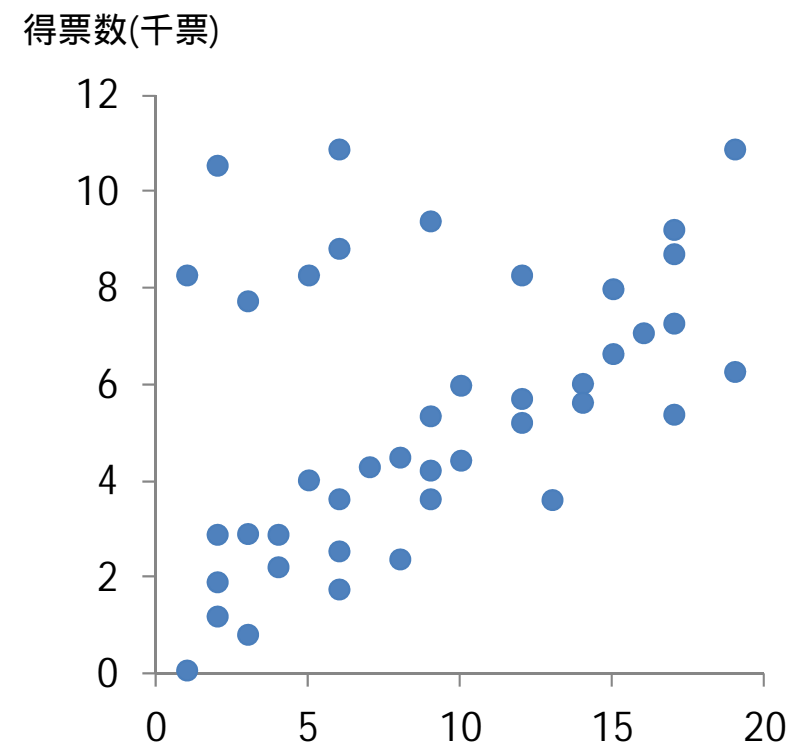
- ✓ 二変数の関連性
(因果、相関)を見たいとき

メリット

- ✓ 二変数の関係性が
目で確認できる
- ✓ 簡単に予測式(回帰式)
が作れる

注意点

- ✓ 相関関係があっても因果
関係があるとは限らない
- ✓ 回帰式に物理的な根拠はない
(統計的根拠のみ持つ)



考える順序と作業手順は逆順になります。想定する結果から順に何が必要かをたどっていきます。

作業手順



考える順序

