

# アメダス気象データ分析 チャレンジ！入門

## 2. データ分析の流れ

主催：気象ビジネス推進コンソーシアム

共催：岐阜大学工学部附属応用気象研究センター

資料作成：吉野 純（岐阜大学）



# 本教材について

**Copyright 2022 気象ビジネス推進コンソーシアム、岐阜大学 吉野純**

**(C) 2022 WXBC、岐阜大学 吉野純**

## <利用条件>

本書は、本書に記載した要件・技術・方式に関する内容が変更されないこと、および出典を明示いただくことを前提に、無償でその全部または一部を複製、翻案、翻訳、転記、引用、公衆送信等して利用できます。なお、全体を複製、翻案、翻訳された場合は、本書にある著作権表示および利用条件を明示してください。

## <免責事項>

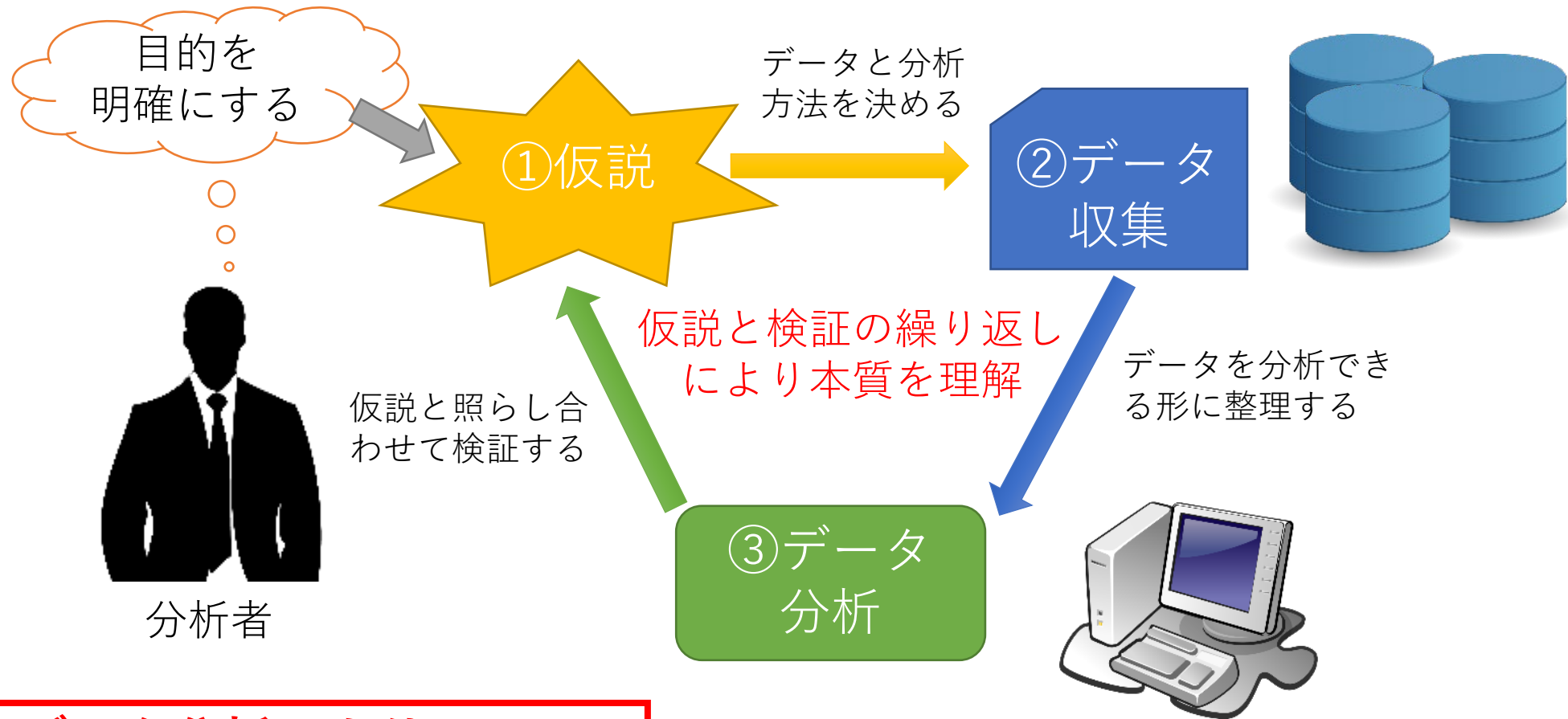
本書の著作権者は、本書の記載内容に関して、その正確性、商品性、利用目的への適合性等に関して保証するものではなく、特許権、著作権、その他の権利を侵害していないことを保証するものでもありません。本書の利用により生じた損害について、本書の著作権者は、法律上のいかなる責任も負いません。

# まずは目的を明確にする

- 「そこにデータがあるから」データ分析をするのではなく、何らかの目的があり、その目的を達成するためにデータ分析をする。まずは、目的を言葉で表現してみよう。
  - （目的）スポーツドリンクの最適な販促計画を立てたい
    - （仮説）気温と売上との間を比較して特徴を見つける
  - （目的）スタジアムの最適な人員配置計画を立てたい
    - （仮説）天気と来場者数との間を比較して特徴を見つける
  - （目的）自動車販売の地域にあった営業計画を立てたい
    - （仮説）降雪量と四輪駆動車の販売数との関係性を比較して特徴を見つける

仮説が正しいかデータで検証する必要がある！

# 仮説と検証の繰り返し



## データ分析により・・・

- ① ビジネス現場で、データに基づいた適切な意思決定ができるようになります。
- ② データ分析の繰り返しにより、分析者の知識や経験が増えて、データ分析の精度が高くなります。

# データ分析の流れ

## □ データの入手

仮説を検証するのに必要なデータを入手する。

▶ 適切なデータを入手できるかが分析のカギ。

## □ データの加工・整理・要約

分析に必要な情報を並び替えや集計によってデータを要約する。

▶ 全体的な傾向や特徴が分かる。

## □ データの可視化

データの傾向や特徴を視覚的に分かりやすくグラフに表す。

▶ 数字では分かりにくかった特徴の発見に役立つ。

## □ データの分析

統計関数を利用してデータの傾向や特徴を数値で示す。

▶ 直感的な理解に根拠を与える。説得力が増す！

# データの入手

## さまざまなオープンデータ

- **政府統計** : <https://www.e-stat.go.jp/>
- RESAS : <https://resas.go.jp/#/47/47201>
- 総務省 : <https://www.data.go.jp/>
- 地方自治体 : <https://www.data.go.jp/list-of-database/local-government/>
- 東京都 : <https://portal.data.metro.tokyo.lg.jp/>
- **気象庁** : <https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
- **東京電力** : <https://www.tepco.co.jp/forecast/>
- Scikit-learn datasets: <https://scikit-learn.org/stable/datasets.html>
  - ✓ トイデータセット : アイリスの種類、ボストンの住宅価格、手書き文字、等
  - ✓ 実世界データセット : 顔画像、ニュース記事、木の種類、カリフォルニアの住宅価格、等
- StatsModels datasets: <https://www.statsmodels.org/0.8.0/datasets/index.html>
  - ✓ 二酸化炭素濃度、エルニーニョ、犯罪、銅市場、ナイル川流量、等

# データの加工・整理・要約

生データのままでは分析に使えないことが多いため、分析に適した形に加工・整理・要約する。

- 数量化
- 欠測値処理
- 平滑化
- 並び替え
- フィルター
- 検索
- クロス集計

**ピボット  
テーブル**  
(クロス集計)

性別毎の味の評価の平均

	A	B
1		
2		
3	行ラベル	平均 / 味の評価採点
4	女性	6.8
5	男性	5.4
6	総計	6.1
7		

性別毎の好みの具の総計

	A	B	C	D	E
1					
2	好みの具(大根)	列ラベル			
3	行ラベル	ちくわ	はんぺん	大根	総計
4	女性	1	0	4	5
5	男性	2	2	1	5
6	総計	3	2	5	10
7					
8					

**VLOOKUP関数**

fx `=VLOOKUP(H4,A3:E13,2,FALSE)`

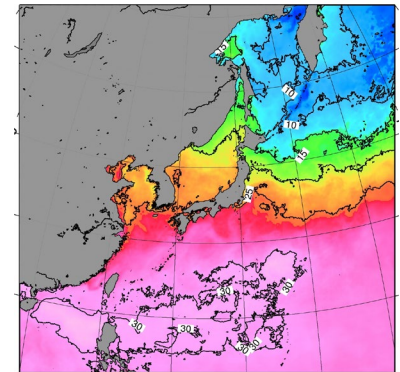
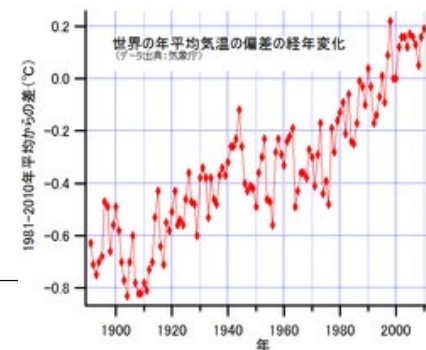
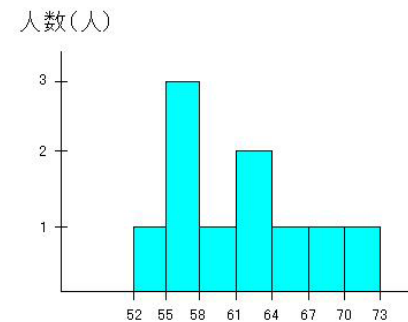
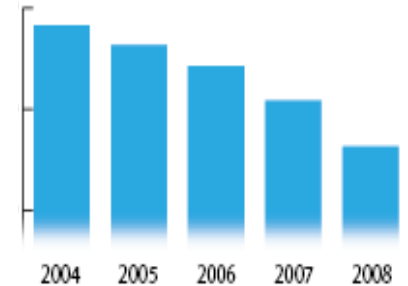
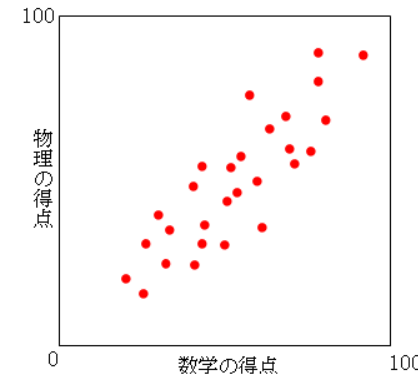
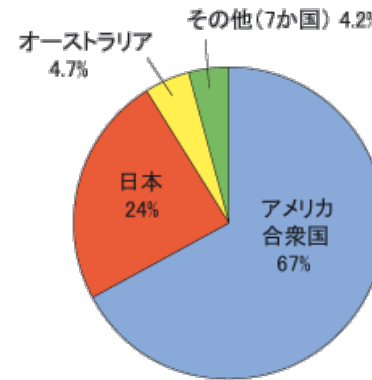
	A	B	C	D	E	F	G	H	I
1	おでんの具アンケートデータ								
2									
3	No.	年齢	性別	味の評価採点	好みの具(大根、ちくわ、はんぺん)			No.	年齢
4	1	26	女性	7	大根			1	26
5	2	50	男性	4	ちくわ			3	15
6	3	15	女性	6	ちくわ			8	19
7	4	20	男性	3	はんぺん			9	30
8	5	28	男性	8	大根				
9	6	19	女性	8	大根				
10	7	33	女性	4	大根				
11	8	19	男性	7	ちくわ				
12	9	30	女性	9	大根				
13	10	35	男性	5	はんぺん				
14									

左のデータベースから  
一部の人情報を抽出できる

# データの可視化

データを適切な形にグラフ化すると、その特徴がみえてくる。数値だけでは分かりにくい例外的な特徴を抽出することもできる。分析の目的に合ったグラフを選ぶよう注意する。

- 円グラフ、パレート図
- 散布図
- 棒グラフ
- ヒストグラム、箱ひげ図
- 折れ線グラフ
- 等値線図、分布図





# データの分析



Gartner Glossaryより <https://www.gartner.com/en/information-technology/glossary/>

- 記述的分析 (descriptive analytics)
  - ▶ 何が起こったのかを可視化により見える化する
- 診断的分析 (diagnostic analytics)
  - ▶ なぜそれが起こったのかを相関関係を調べて明らかにする
- 予測的分析 (predictive analytics)
  - ▶ 予測モデルを作成して将来予測をしながら未来の状態を分析する
- 処方的分析 (prescriptive analytics)
  - ▶ 予測に基づいて最適な行動 (意思決定) をとるための分析

過去の状態をひもとき 未来の状態を知る 予めできる対策を考える

# データの分析

データの中から全体像や特徴を見きわめるために、さまざまな分析手法を利用する。

## □教師なし学習

- ✓相関分析
- ✓主成分分析、因子分析、コレスポンデンス分析
- ✓クラスター分析

・・・ 診断的分析

## □教師あり学習（分類、回帰）

- ✓重回帰分析、ロツソ・ラツソ回帰、ロジスティック回帰
- ✓決定木分析、ランダムフォレスト
- ✓ニューラルネットワーク、SVM、CNN

・・・ 予測的分析

## □時系列データ解析

- ✓SARIM、カルマンフィルタ
- ✓LSTM-RNN

・・・ 予測的分析

## □意思決定分析

- ✓ベイズ推定、アソシエーション分析、決定分析

・・・ 処方的分析

# データの分析

データの中から全体像や特徴を見きわめるために、さまざまな分析手法を利用する。

## Excelの様々な統計関数

目的	関数名	分析例
データを要約する	FREQUENCY	度数分布表を作る
集団の代表値を求める	AVERAGE	平均値を求める
	MEDIAN	中央値を求める
	MODE.SNGL MODE.MULT	最頻値を求める
データのばらつきを知る	STDEV.P	標本標準偏差を求める
全体の中での位置を知る	NORM.DIST	正規分布の累積確率を求める
	PERCENTILE.INC	パーセンタイルを求める
	QUARTILE.INC	四分位数を求める

目的	関数名	分析例
2つの変数の関係を知る	CORREL	相関係数を求める
将来の値を予測する	SLOPE	回帰直線の係数を求める
	INTERCEPT	回帰直線の定数項を求める
	FORECAST.LINEAR	回帰分析を行う
	LINEST	重回帰分析を行う
	FORECAST.ETS	時系列分析を行う
	FORECAST.EST.SEASONALITY	時系列分析の季節性（周期性）を求める