

気象データ活用によるデータ分析の流れ ～ データや分析手法の選び方 ～

岐阜大学工学部附属応用気象研究センター センター長・准教授

気象ビジネス推進コンソーシアム人材育成WG メンバー

気象データ×IT勉強会 リーダー

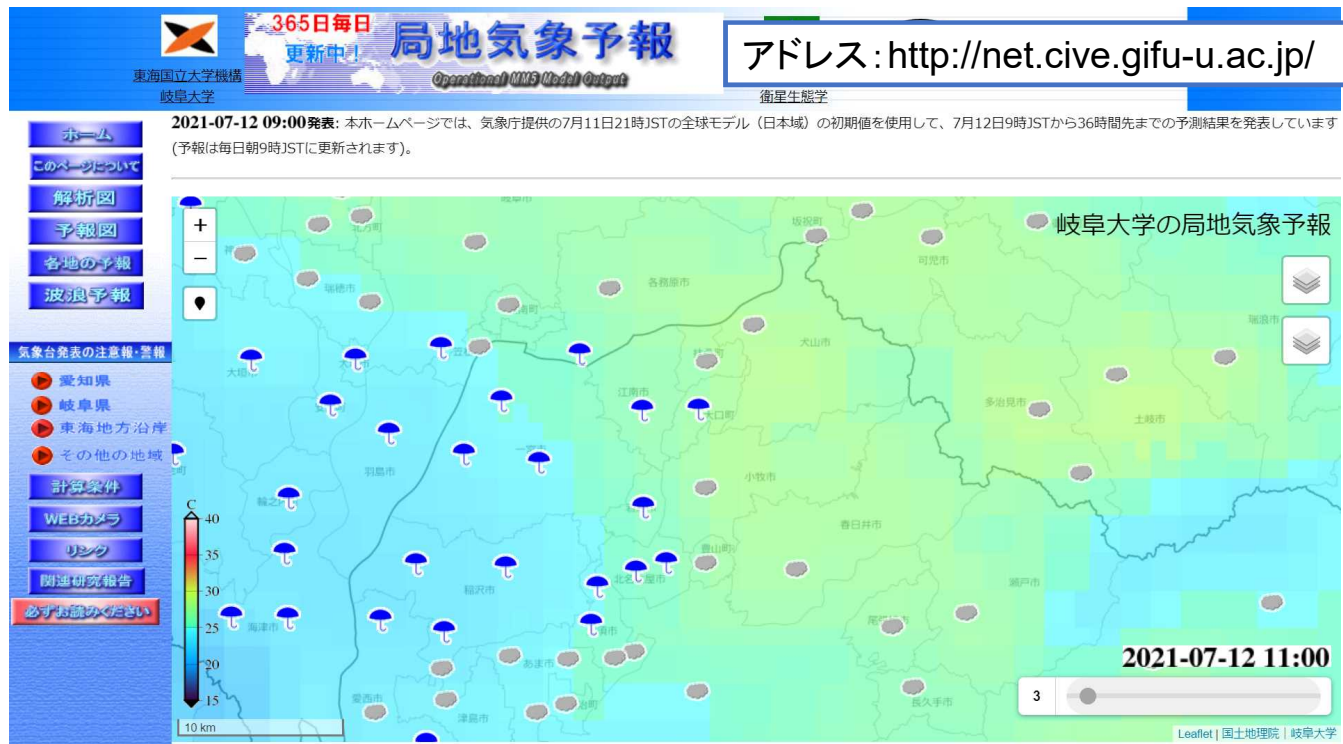
気象データアナリスト「スキルセット検討委員会」 委員長

吉野 純



自己紹介

吉野 純 博士（理学）・気象予報士2220号
岐阜大学工学部社会基盤工学科 准教授
工学部附属応用気象研究センター センター長



【特徴】

- ・大学初で唯一の気象予報業務許可（気象・波浪）の取得
- ・2005年6月1日より予報業務開始
- ・ホームページのみでの情報公開
- ・愛知県、岐阜県を2kmメッシュで予測
- ・1日1回の情報更新（朝9時）
- ・翌日夜9時までの36時間予報

講演内容

- 気象データの分析の流れ
- 気象×〇〇の分析例
- 様々な気象データ
- 様々な分析手法
- 気象データ分析に挑戦してみませんか？

内容：気象データを使ってデータ分析してみたい！そんなときに注意すべきポイントとは？ WXBC人材育成WGが企画する「気象データ分析チャレンジ！」の取り組みを例に気象データ活用によるデータ分析の流れを説明するとともに、様々な気象データや様々な分析手法の中からどう使い分けていくのか解説します。みなさんのビジネスに気象データをもっと活用してみませんか？

講演内容

- 気象データの分析の流れ
- 気象×〇〇の分析例
- 様々な気象データ
- 様々な分析手法
- 気象データ分析に挑戦してみませんか？

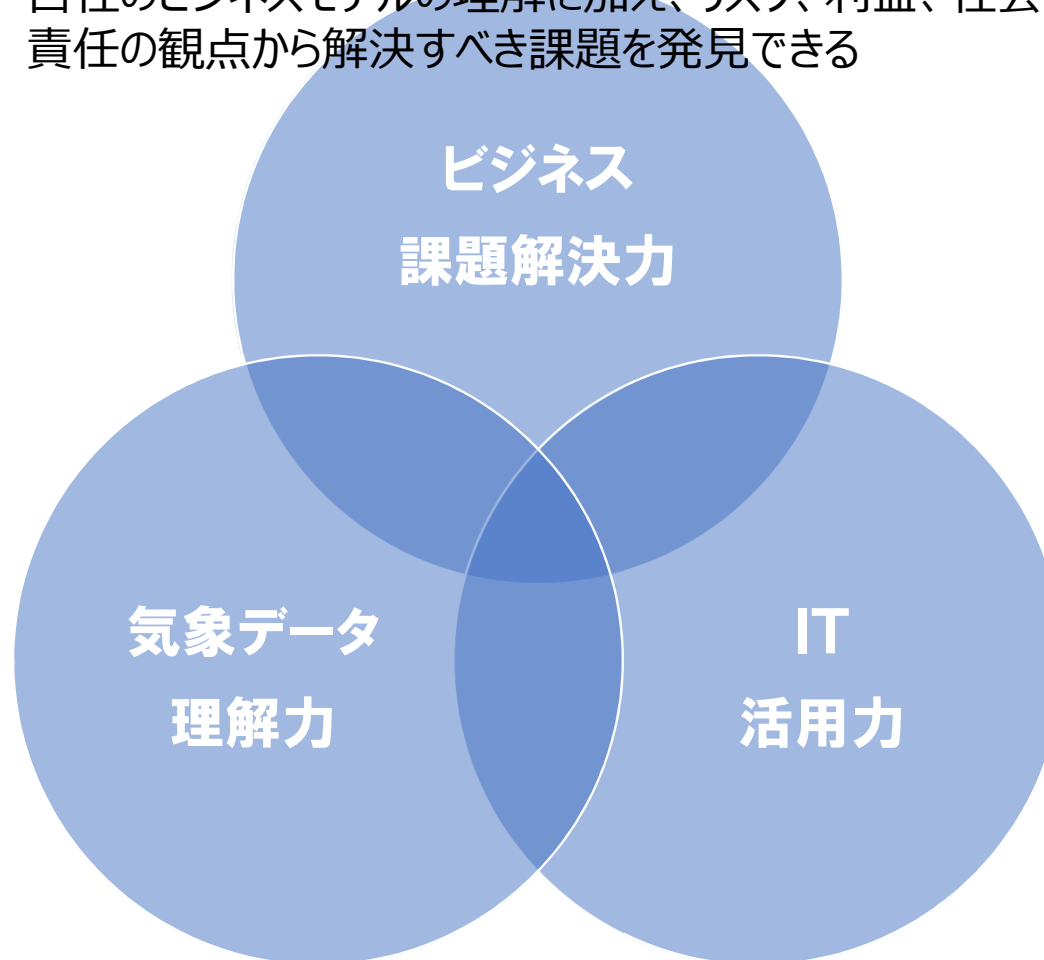
内容：気象データを使ってデータ分析してみたい！そんなときに注意すべきポイントとは？ WXBC人材育成WGが企画する「気象データ分析チャレンジ！」の取り組みを例に気象データ活用によるデータ分析の流れを説明するとともに、様々な気象データや様々な分析手法の中からどう使い分けていくのか解説します。みなさんのビジネスに気象データをもっと活用してみませんか？

気象データ分析に必要な能力



気象データアナリスト育成講座 カリキュラムガイドラインより

自社のビジネスモデルの理解に加え、リスク、利益、社会的責任の観点から解決すべき課題を発見できる



気象データの特徴を理解した上で、課題解決のための仮説を立てられるようになる

データサイエンスのスキルを使って仮説を検証し、課題解決ができるようになる

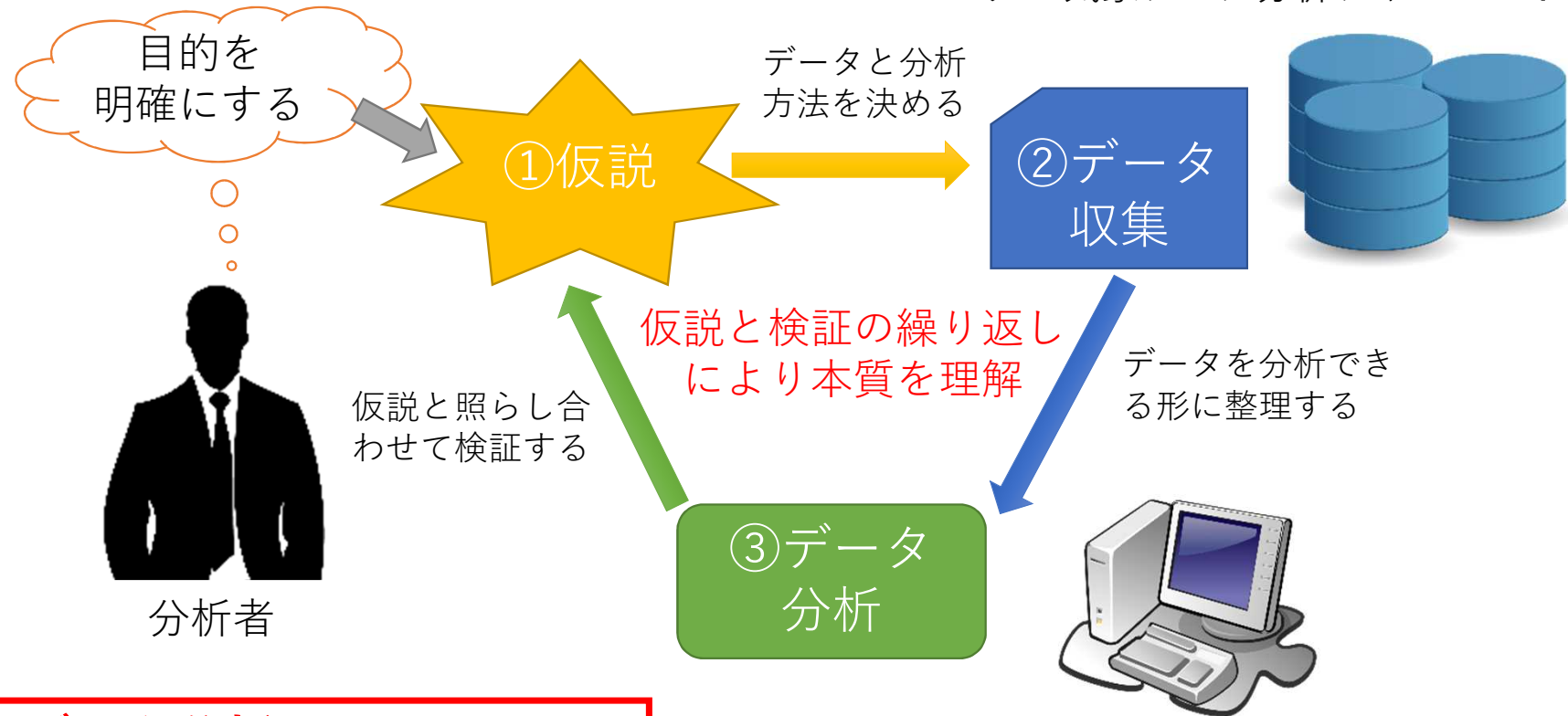
まずは目的を明確にする

- 「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

アメダス気象データ分析チャレンジ！(Excel版)より



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手に入る。

➤適切なデータを手に入るかが分析のカギ。

□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

アメダス気象データ分析チャレンジ！（Excel版）より

さまざまなオープンデータ

- 政府統計：<https://www.e-stat.go.jp/>
- 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

アメダス気象データ分析チャレンジ！（Excel版）より

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

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

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	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
3	1	26	女性	7	大根			1	26
4	2	50	男性	4	ちくわ			3	15
5	3	15	女性	6	ちくわ			8	19
6	4	20	男性	3	はんぺん			9	30
7	5	28	男性	8	大根				
8	6	19	女性	8	大根				
9	7	33	女性	4	大根				
10	8	19	男性	7	ちくわ				
11	9	30	女性	9	大根				
12	10	35	男性	5	はんぺん				

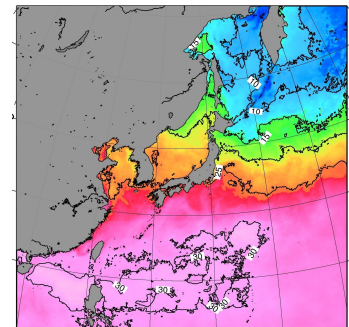
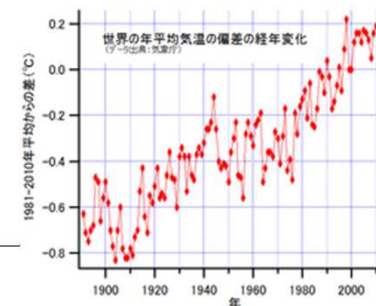
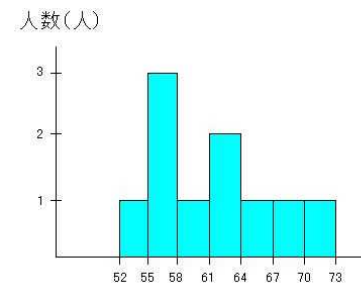
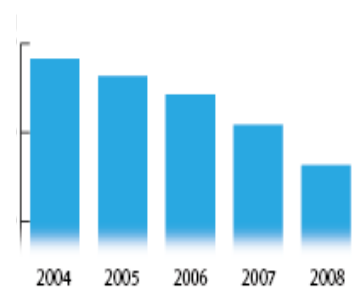
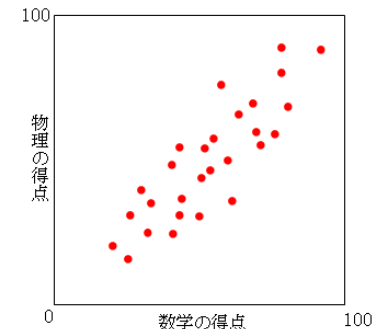
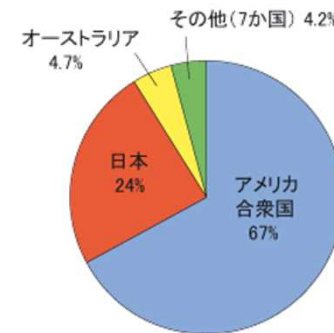
左のデータベースから
一部の人情報を抽出できる

データの可視化

アメダス気象データ分析チャレンジ！（Excel版）より

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

アメダス気象データ分析チャレンジ！（Excel版）より

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）

- ✓回帰分析、重回帰分析、決定木分析
- ✓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

- ✓ARMA、ARIMA、SARIM、カルマンフィルタ
- ✓LSTM-RNN

．．．予測的分析

□意思決定分析

- ✓決定木分析、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 FORECAST.ETS.SEASONALITY	時系列分析の季節性（周期性）を求める
	FORECAST.ETS	時系列分析を行う

講演内容

- 気象データの分析の流れ
- 気象×〇〇の分析例
- 様々な気象データ
- 様々な分析手法
- 気象データ分析に挑戦してみませんか？

内容：気象データを使ってデータ分析してみたい！そんなときに注意すべきポイントとは？ WXBC人材育成WGが企画する「気象データ分析チャレンジ！」の取り組みを例に気象データ活用によるデータ分析の流れを説明するとともに、様々な気象データや様々な分析手法の中からどう使い分けていくのか解説します。みなさんのビジネスに気象データをもっと活用してみませんか？

データ分析してみよう

アメダス気象データ分析チャレンジ！（Excel版）より

例えば、

	A	B	C	D	E
1	気温とおでんの売上個数の関係				
2	日付	気温	販売数		
3	10月1日	20.9	22		
4	10月2日	22.3	32		
5	10月3日	22.1	19		
6	10月4日	18.3	22		
7	10月5日	16.9	30		
8	10月6日	15.7	37		
9	10月7日	18.0	36		
10	10月8日	20.9	22		
11	10月9日	21.0	59		
12	10月10日	22.8	20		
13	10月11日	21.2	21		
14	10月12日	23.1	19		
15	10月13日	15.9	15		

A列に日づけ

B列に気温が入
力されている

C列に販売数が入
力されている

（注意1）実際には
これほどよく整理さ
れたデータがすぐに
手に入るわけではあ
りません！データの
整理・加工が必要に
なることがほとんど
です。

（注意2）このデー
タは人工的なもので
現実を反映していま
せん。

（仮説）おでんの販売数と気温との間には関係性があるのではないか？

エクセルで時系列図

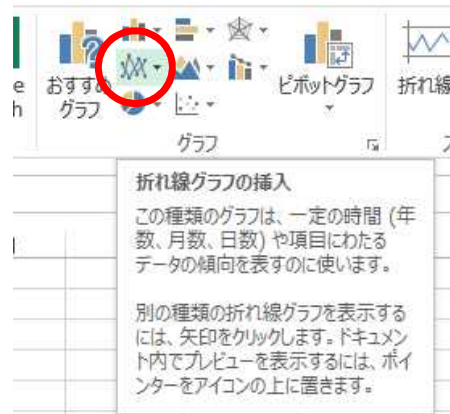
アメダス気象データ分析チャレンジ！（Excel版）より

	A	B	C	D	E
1	気温とおでんの売上個数の関係				
2	日付	気温	販売数		
3	10月1日	20.9	22		
4	10月2日	22.3	32		
5	10月3日	22.1	19		
6	10月4日	18.3	22		
7	10月5日	16.9	30		
8	10月6日	15.7	37		
9	10月7日	18.0	36		
10	10月8日	20.9	22		
11	10月9日	21.0	59		
12	10月10日	22.8	20		
13	10月11日	21.2	21		
14	10月12日	23.1	19		
15	10月13日	15.9	15		

A列とB列を選択する

	A	B	C	D	E
1	気温とおでんの売上個数の関係				
2	日付	気温	販売数		
3	10月1日	20.9	22		
4	10月2日	22.3	32		
5	10月3日	22.1	19		
6	10月4日	18.3	22		
7	10月5日	16.9	30		
8	10月6日	15.7	37		
9	10月7日	18.0	36		
10	10月8日	20.9	22		
11	10月9日	21.0	59		
12	10月10日	22.8	20		
13	10月11日	21.2	21		
14	10月12日	23.1	19		
15	10月13日	15.9	15		

A列とC列を選択する（Ctrlを押しながら）



ツールバーの中から
折れ線グラフをクリック



気温の時系列図が描けた



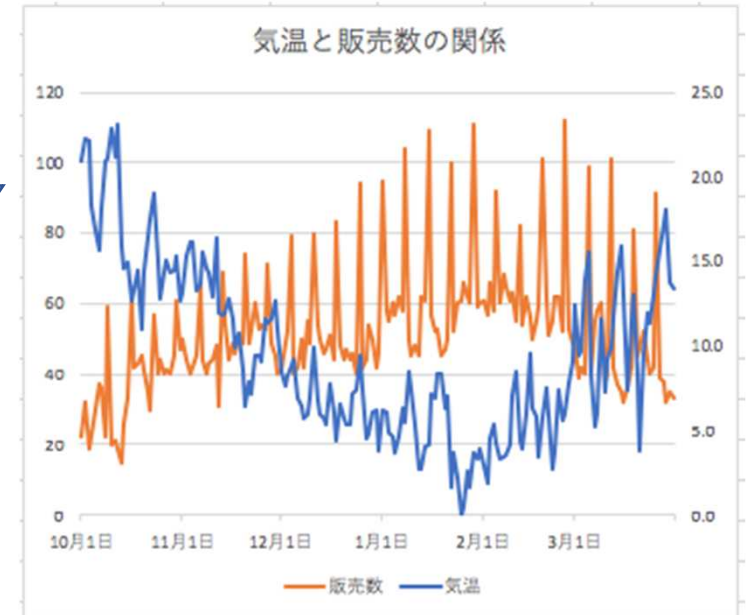
販売数の時系列図が描けた

エクセルで時系列図

アメダス気象データ分析チャレンジ！（Excel版）より

	A	B	C	D	E
1	気温とおでんの売上個数の関係				
2	日付	気温	販売数		
3	10月1日	20.9	22		
4	10月2日	22.3	32		
5	10月3日	22.1	19		
6	10月4日	18.3	22		
7	10月5日	16.9	30		
8	10月6日	15.7	37		
9	10月7日	18.0	36		
10	10月8日	20.9	22		
11	10月9日	21.0	59		
12	10月10日	22.8	20		
13	10月11日	21.2	21		
14	10月12日	23.1	19		
15	10月13日	15.9	15		

A列からC列まで選択する



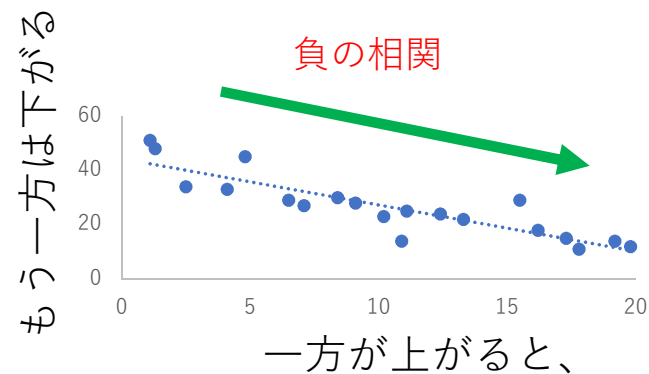
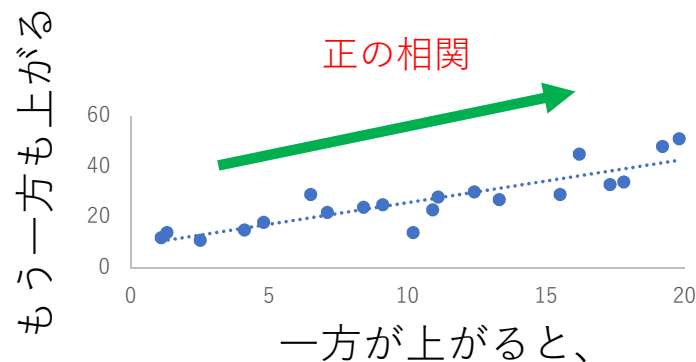
気温と販売数を重ねた時系列図が描けた

ここからぼんやりと読み取れること：

- 秋からの冬へと気温が下がるにつれて販売数が増える。
- 週に1回、異常によく売れている日がある。日曜特売？
- 気温が上がった翌日には販売数が減る傾向にある。

相関関係

- 時系列図から、気温とおでんの販売数の間から関係性を想像することができました。「一方が上がれば、もう一方も上がる（**正の相関**）」という関係や「一方が上がれば、もう一方は下がる（**負の相関**）」といった関係を「**相関関係がある**」（あるいは単に「**相関がある**」）と言います。
- 先の時系列図から「気温とおでんの売れ行きには負の相関がある」と推測されます。



エクセルで散布図

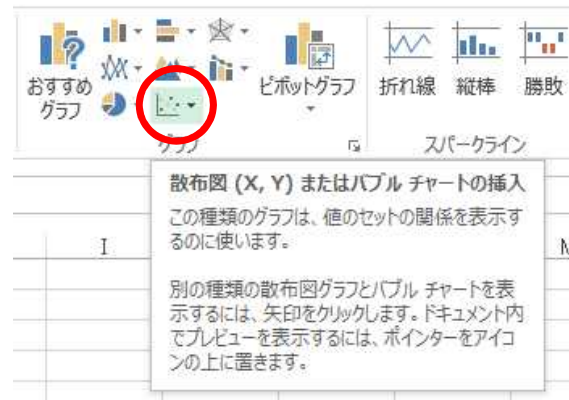
アメダス気象データ分析チャレンジ！（Excel版）より

- 相関があるかどうかを確かめたいときには、まず、散布図を作成します。

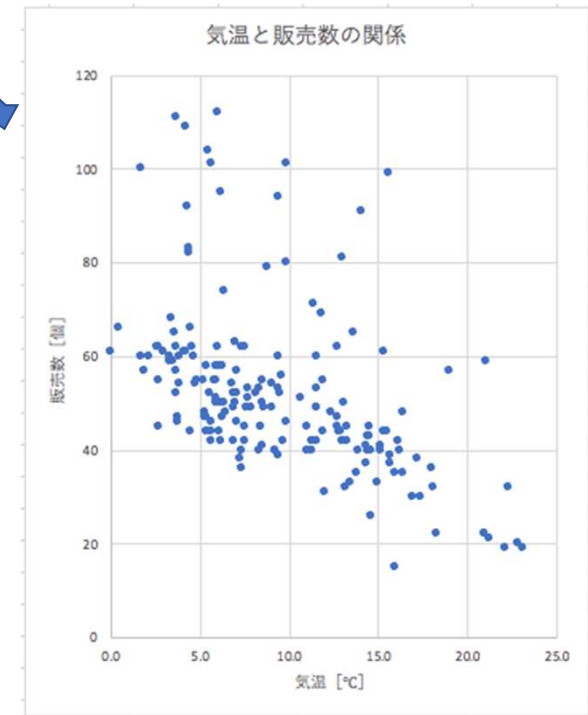
右肩さがり！負の相関がありそうだ！

	A	B	C	D	E
1	気温とおでこの売上個数の関係				
2	日付	気温	販売数		
3	10月1日	20.9	22		
4	10月2日	22.3	32		
5	10月3日	22.1	19		
6	10月4日	18.3	22		
7	10月5日	16.9	30		
8	10月6日	15.7	37		
9	10月7日	18.0	36		
10	10月8日	20.9	22		
11	10月9日	21.0	59		
12	10月10日	22.8	20		
13	10月11日	21.2	21		
14	10月12日	23.1	19		
15	10月13日	15.9	15		

B列とC列を選択する



ツールバーの中から
散布図をクリック

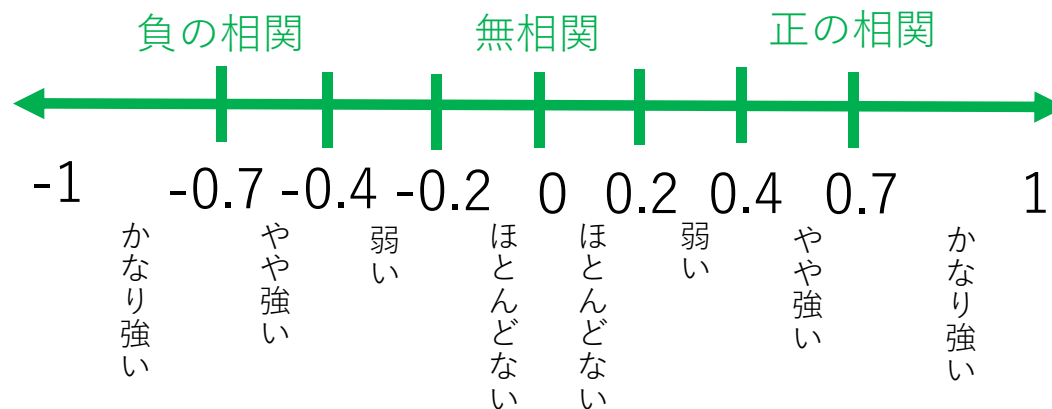


気温と販売数の散布図を描けた

相関係数

アメダス気象データ分析チャレンジ！(Excel版)より

- 散布図により負の相関があることが読み取れたが、それを数字などで示せないと他の人にはなかなか納得してもらえません。
- 相関関係の強さを表す指標に、「**相関係数**」があり、 $-1 \sim 1$ の数字で表します。相関係数が1に近いほど正の相関が強く、相関係数が -1 に近いほど負の相関が強くなります。

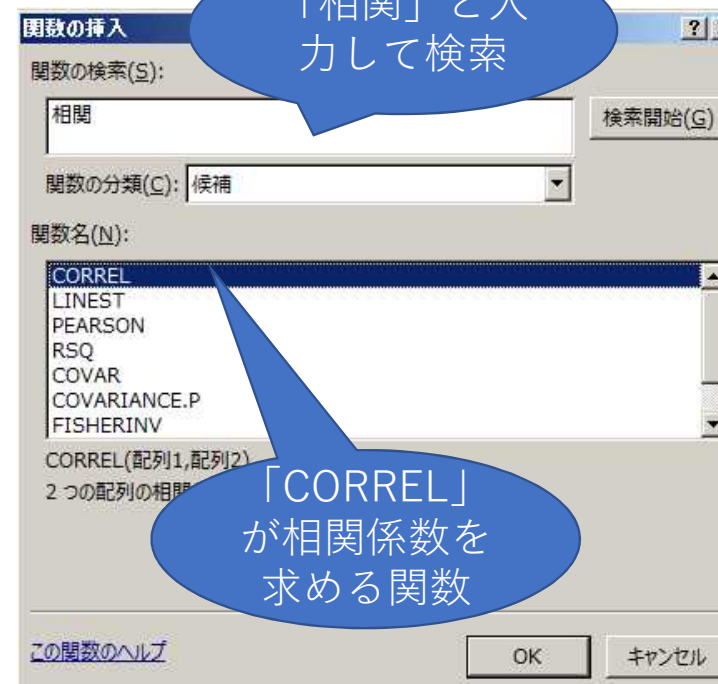


エクセルで相関係数

アメダス気象データ分析チャレンジ！(Excel版) より

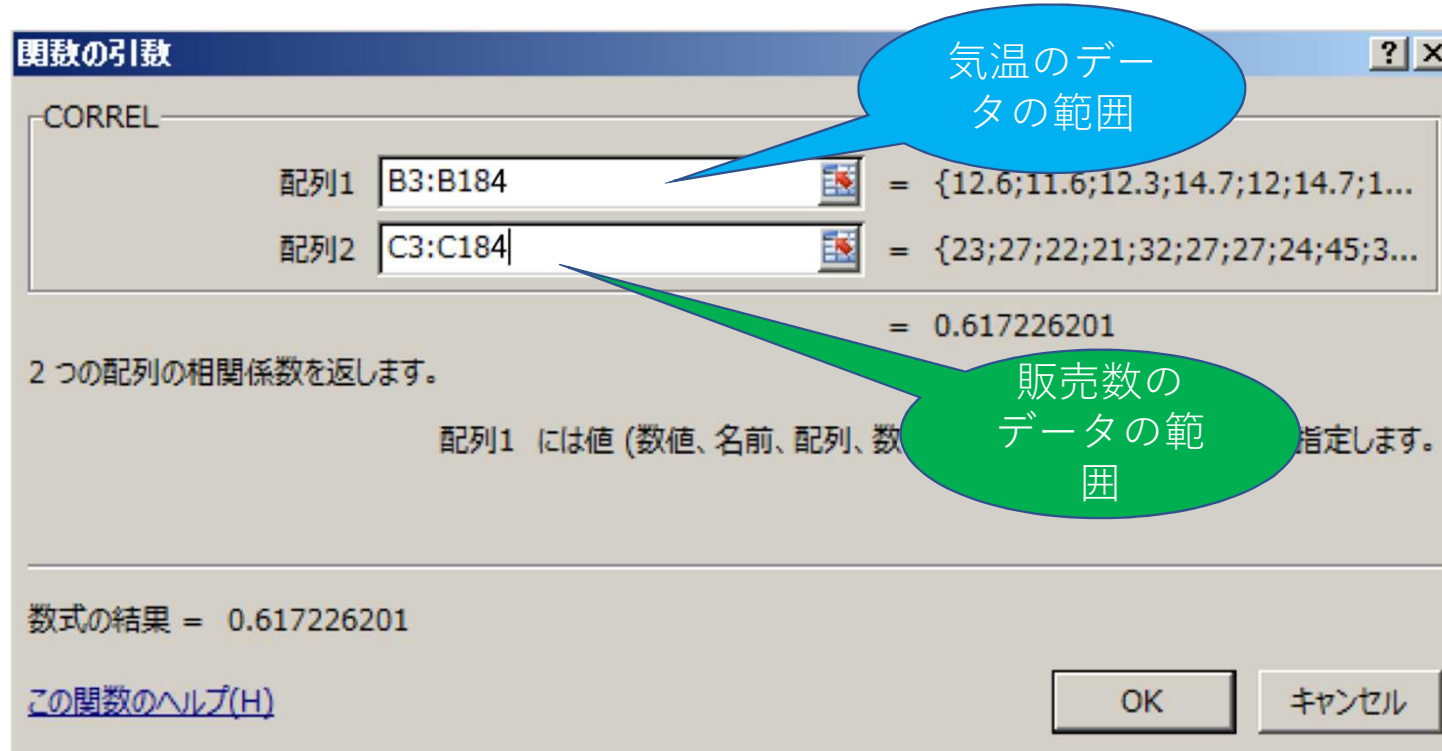
- 相関の強さを表す相関係数は、エクセルの関数「CORREL」で計算できます。

	ホーム	挿入	描画	ページレイアウト	数式	データ
						
	関数の挿入	オートSUM	最近使ったもの	財務	論理	テキスト
D3						
	A	B	C	D	E	F
1	気温とおでんの売上個数の関係					
2	日付	気温	販売数	相関係数		
3	10月1日	20.9	22			
4	10月2日	22.3	32			
5	10月3日	22.1	19			
6	10月4日	18.3	22			
7	10月5日	16.9	30			
8	10月6日	15.7	37			
9	10月7日	18.0	36			
10	10月8日	20.9	22			



エクセルで相関係数

アメダス気象データ分析チャレンジ！（Excel版）より



または、セルの中で関数とデータの配列を直接打ち込む。

=CORREL(B3:B184,C3:C184)

気温の範囲 販売量の範囲

エクセルで相関係数

アメダス気象データ分析チャレンジ！（Excel版）より

D3				=CORREL(B3:B184,C3:C184)			
	A	B	C	D	E	F	G
1	気温とおでんの売上個数の関係						
2	日付	気温	販売数	相関係数			
3	10月1日	20.9	22	-0.50483			
4	10月2日	22.3	32				
5	10月3日	22.1	19				
6	10月4日	18.3	22				
7	10月5日	16.9	30				
8	10月6日	15.7	37				
9	10月7日	18.0	36				
10	10月8日	20.9	22				
11	10月9日	21.0	59				
12	10月10日	22.8	20				
13	10月11日	21.2	21				

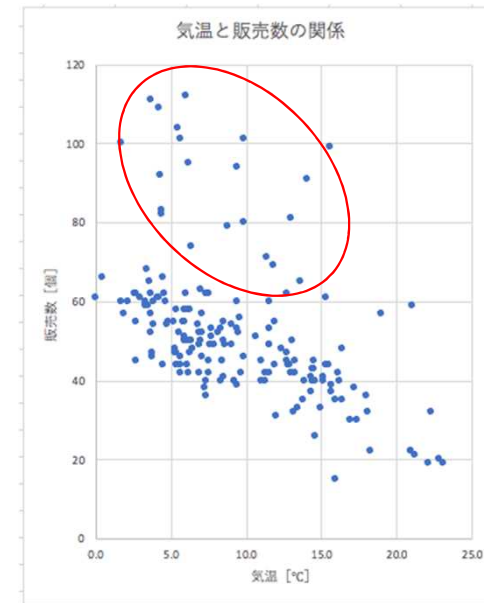
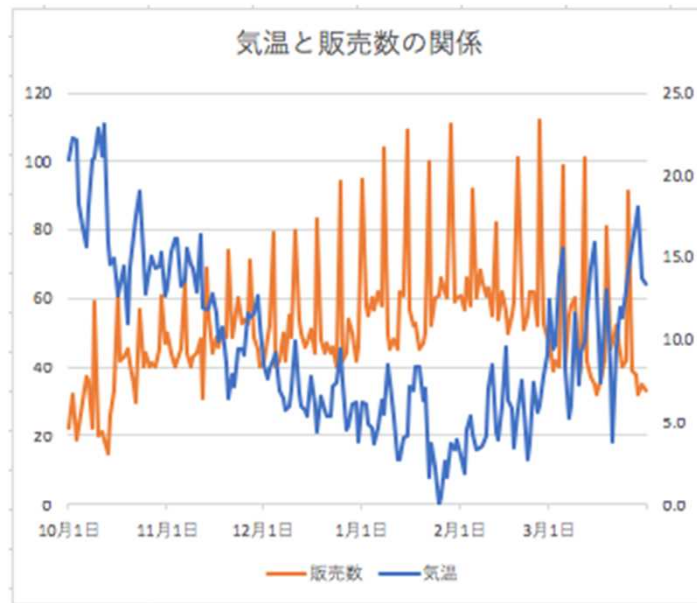
やや強い負の
相関がある

- 気温とおでんの販売数の関係の間にはやや強い相関があることが数値で示すことができた！（仮説が正しいことが示された）

ここで再び仮説！

アメダス気象データ分析チャレンジ！（Excel版）より

- 気温と販売数の間にはもっと強い相関関係があると想像されましたが、相関係数は0.50と「やや強い」程度にとどまりました。これは何故でしょう？
- もしかしたら日曜特売の異常なデータが悪さをしているのではないだろうか？これらを取り除いたら、気温と販売数の間の相関関係はより強くなるだろうか？



データの整理

アメダス気象データ分析チャレンジ！（Excel版）より

- まず、日曜日のデータを空文字「」にする。以下のように、IF関数やWEEKDAY関数を利用する。

=IF(WEEKDAY(\$A3,1)=1,"",B3)

E3						
	A	B	C	D	E	F
1	気温とおでんの売上個数の関係					
2	日付	気温	販売数	相関係数	気温（除日曜）	販売数（除日曜）
3	10月1日	20.9	22	-0.50483	20.9	
4	10月2日	22.3	32			
5	10月3日	22.1	19			
6	10月4日	18.3	22			
7	10月5日	16.9	30			
8	10月6日	15.7	37			
9	10月7日	18.0	36			
10	10月8日	20.9	22			
11	10月9日	21.0	59			
12	10月10日	22.8	20			
13	10月11日	21.2	21			
14	10月12日	23.1	19			
15	10月13日	15.9	15			
16	10月14日	14.6	26			

日付（A3）が日曜日（=1）ならば空文字「」、日曜日でない（≠1）ならば気温（B3）をそのまま置きなさいという意味

入力後、ここをドラッグして、数式をその他のセルにコピーする

データの整理

アメダス気象データ分析チャレンジ！（Excel版）より

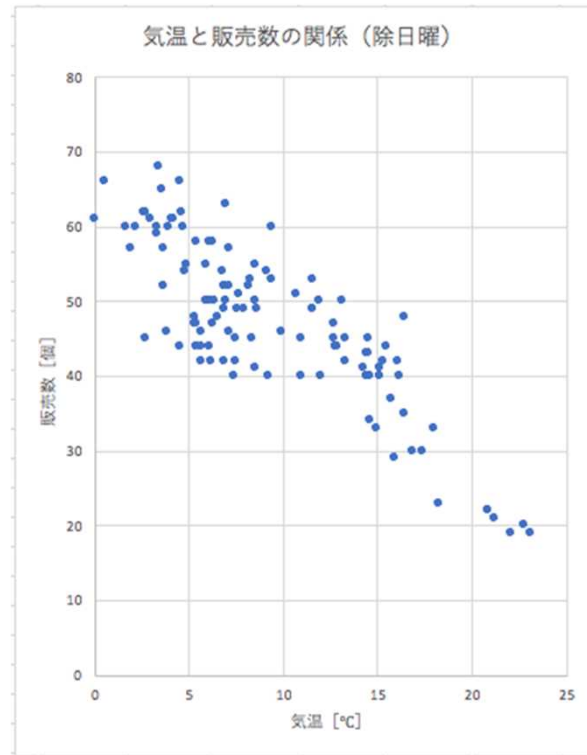
- 日曜日を取り除いたデータが作成できた。同様に、散布図や相関係数で検証してみよう。

	A	B	C	D	E	F
1	気温とおでんの売上個数の関係					
2	日付	気温	販売数	相関係数	気温（除日曜）	販売数（除日曜）
3	10月1日	20.9	22	-0.50483	20.9	22
4	10月2日	22.3	32			
5	10月3日	22.1	19		22.1	19
6	10月4日	18.3	22		18.3	22
7	10月5日	16.9	30		16.9	30
8	10月6日	15.7	37		15.7	37
9	10月7日	18.0	36		18	36
10	10月8日	20.9	22		20.9	22
11	10月9日	21.0	59			
12	10月10日	22.8	20		22.8	20
13	10月11日	21.2	21		21.2	21
14	10月12日	23.1	19		23.1	19
15	10月13日	15.9	15		15.9	15
16	10月14日	14.6	26		14.6	26
17	10月15日	15.0	33		15	33
18	10月16日	12.7	62			
19	10月17日	13.3	42		13.3	42
20	10月18日	14.5	43		14.5	43
21	10月19日	11.0	45		11	45
22	10月20日	14.3	41		14.3	41
23	10月21日	16.4	35		16.4	35
24	10月22日	17.4	30		17.4	30
25	10月23日	19.0	57			
26	10月24日	15.1	40		15.1	40

日曜日のデータ
は空白となった

データの再検証

アメダス気象データ分析チャレンジ！（Excel版）より



より強い相関関係が確認できる

E	F	G
気温（除日曜）	販売数（除日曜）	相関係数（除日曜）
20.9	22	-0.794598664
22.1	19	
18.3	22	
16.9	30	
15.7	37	
18	36	
20.9	22	
22.8	20	
21.2	21	
23.1	19	
15.9	15	
14.6	26	
15	33	

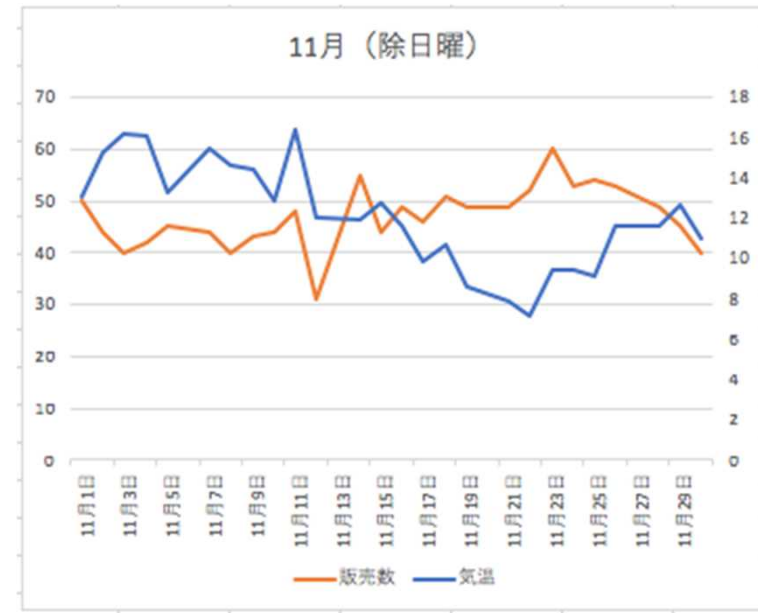
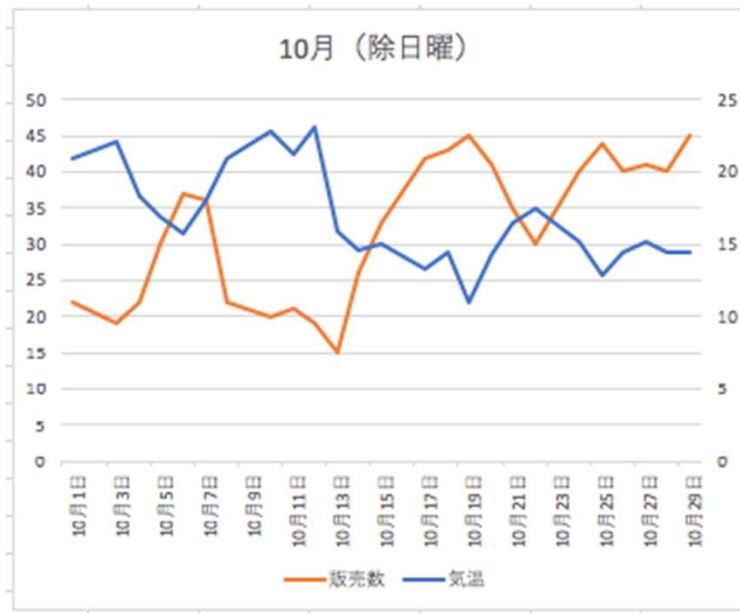
相関係数は0.794と「強い負の相関」

- 気温と日曜以外の販売数の間には強い負の相関があることが明らかとなった。（仮説が正しいことが示された）

ここでまたまた仮説！

アメダス気象データ分析チャレンジ！（Excel版）より

- おでんは前日に気温が下がると販売数が上がる傾向にあり、当日よりも前日の気温との相関関係の方が強いことが想像される。時系列図を拡大することでも、その傾向が確認された。
- **そこで、日曜を除いた販売量を1日前にずらして気温と比較することにより、相関関係はより強くなるのではないのだろうか？**



データの整理

アメダス気象データ分析チャレンジ！（Excel版）より

- まず、10月1日のセル（H3）に一日後の販売数（F4）をそのまま置く。

	A	B	C	D	E	F	G	H
1	気温とおでんの売上個数の関係							
2	日付	気温	販売数	相関係数	気温（除日曜）	販売数（除日曜）	相関係数（除日曜）	販売数（除日曜+1日後）
3	10月1日	20.9	22	-0.50483	20.9	22	-0.794598664	=F4
4	10月2日	22.3	32					
5	10月3日	22.1	19		22.1	19		
6	10月4日	18.3	22		18.3	22		
7	10月5日	16.9	30		16.9	30		
8	10月6日	15.7	37		15.7	37		
9	10月7日	18.0	36		18	36		
10	10月8日	20.9	22		20.9	22		
11	10月9日	21.0	59					
12	10月10日	22.8	20		22.8	20		
13	10月11日	21.2	21		21.2	21		
14	10月12日	23.1	19		23.1	19		
15	10月13日	15.9	15		15.9	15		
16	10月14日	14.6	26		14.6	26		

入力後、ここをドラッグして、
数式をその他のセルにコピーする

データの整理

アメダス気象データ分析チャレンジ！（Excel版）より

- 販売数を1日前にずらすことができた。次に、それに対応する気温のデータも作成する。

	A	B	C	D	E	F	G	H
1	気温とおでんの売上個数の関係							
2	日付	気温	販売数	相関係数	気温（除日曜）	販売数（除日曜）	相関係数（除日曜）	販売数（除日曜+1日後）
3	10月1日	20.9	22	-0.50483	20.9	22	-0.794598664	
4	10月2日	22.3	32					19
5	10月3日	22.1	19		22.1	19		22
6	10月4日	18.3	22		18.3	22		30
7	10月5日	16.9	30		16.9	30		37
8	10月6日	15.7	37		15.7	37		36
9	10月7日	18.0	36		18	36		22
10	10月8日	20.9	22		20.9	22		
11	10月9日	21.0	59					20
12	10月10日	22.8	20		22.8	20		21
13	10月11日	21.2	21		21.2	21		19
14	10月12日	23.1	19		23.1	19		15
15	10月13日	15.9	15		15.9	15		26
16	10月14日	14.6	26		14.6	26		33

データの整理

アメダス気象データ分析チャレンジ！（Excel版）より

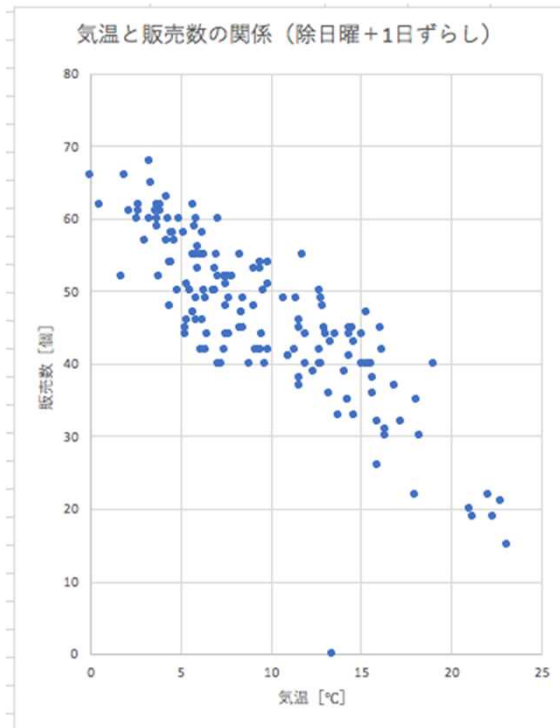
- IF関数を利用して、1日前にずらした販売数（日曜を除く）が空白でない日のみ気温を埋める。

=IF(H3="", "", B3)

fx =IF(H3="", "", B3)							
B	C	D	E	F	G	H	I
人の売上個数の関係							
気温	販売数	相関係数	気温（除日曜）	販売数（除日曜）	相関係数（除日曜）	販売数（除日曜+1日後）	気温（除空白）
20.9	22	-0.50483	20.9	22	-0.794598664		
22.3	32					19	22.3
22.1	19		22.1	19		22	22.1
18.3	22		18.3	22		30	18.3
16.9	30		16.9	30		37	16.9
15.7	37		15.7	37		36	15.7
18.0	36		18	36		22	18
20.9	22		20.9	22			
21.0	59					20	21
22.8	20		22.8	20		21	22.8
21.2	21		21.2	21		19	21.2
23.1	19		23.1	19		15	23.1
15.9	15		15.9	15		26	15.9
14.6	26		14.6	26		33	14.6

データの再々検証

アメダス気象データ分析チャレンジ！（Excel版）より



さらに強い相関関係が確認できる

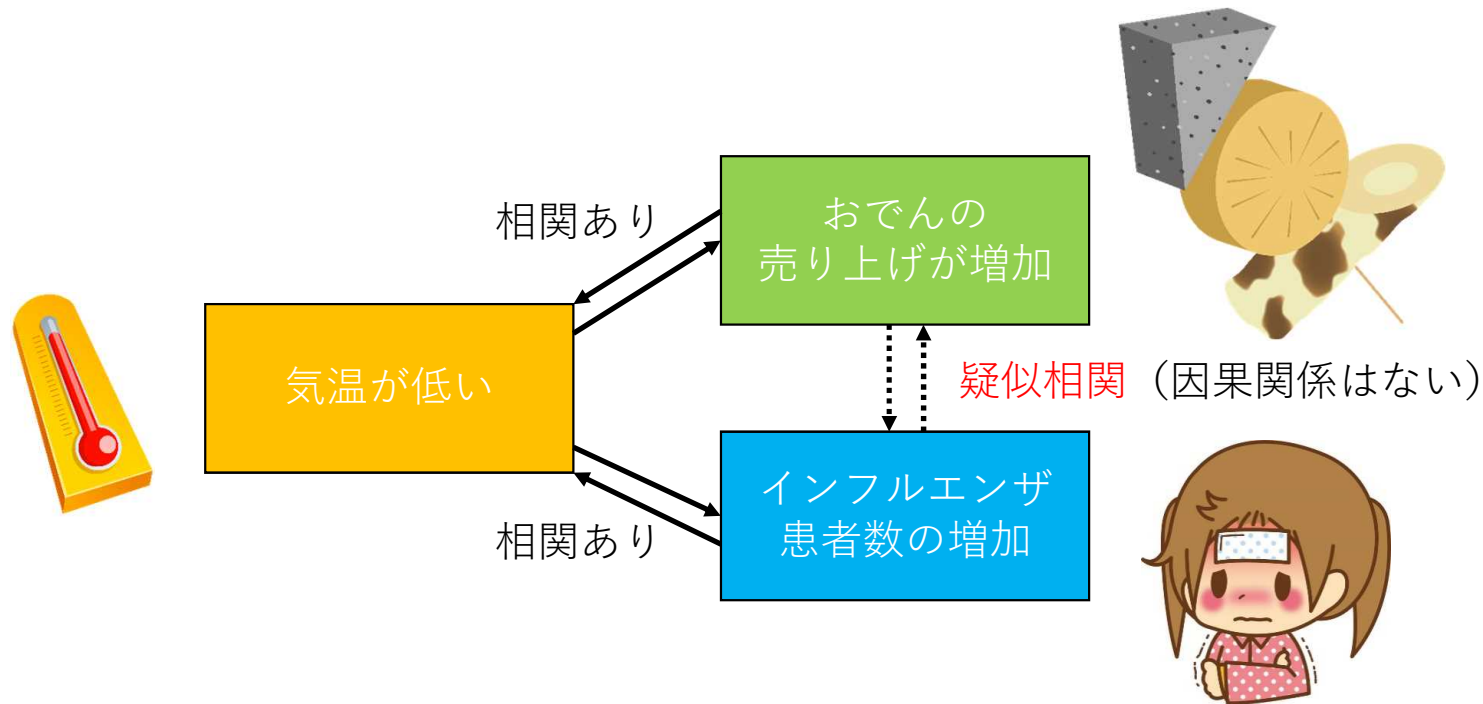
H	I	J
販売数（除日曜+1日後）	気温（除空白）	相関係数（除日曜+1日後）
		-0.831252869
19	22.3	
22	22.1	
30	18.3	
37	16.9	
36	15.7	
22	18	
20	21	
21	22.8	
19	21.2	
15	23.1	
26	15.9	
33	14.6	

相関係数は-0.831と「極めて強い負の相関」

- 気温と日曜以外の販売量（1日ずらし）の間には極めて強い負の相関があることが明らかとなった。（仮説が正しいことが示された）

疑似相関に気をつける

アメダス気象データ分析チャレンジ！（Excel版）より



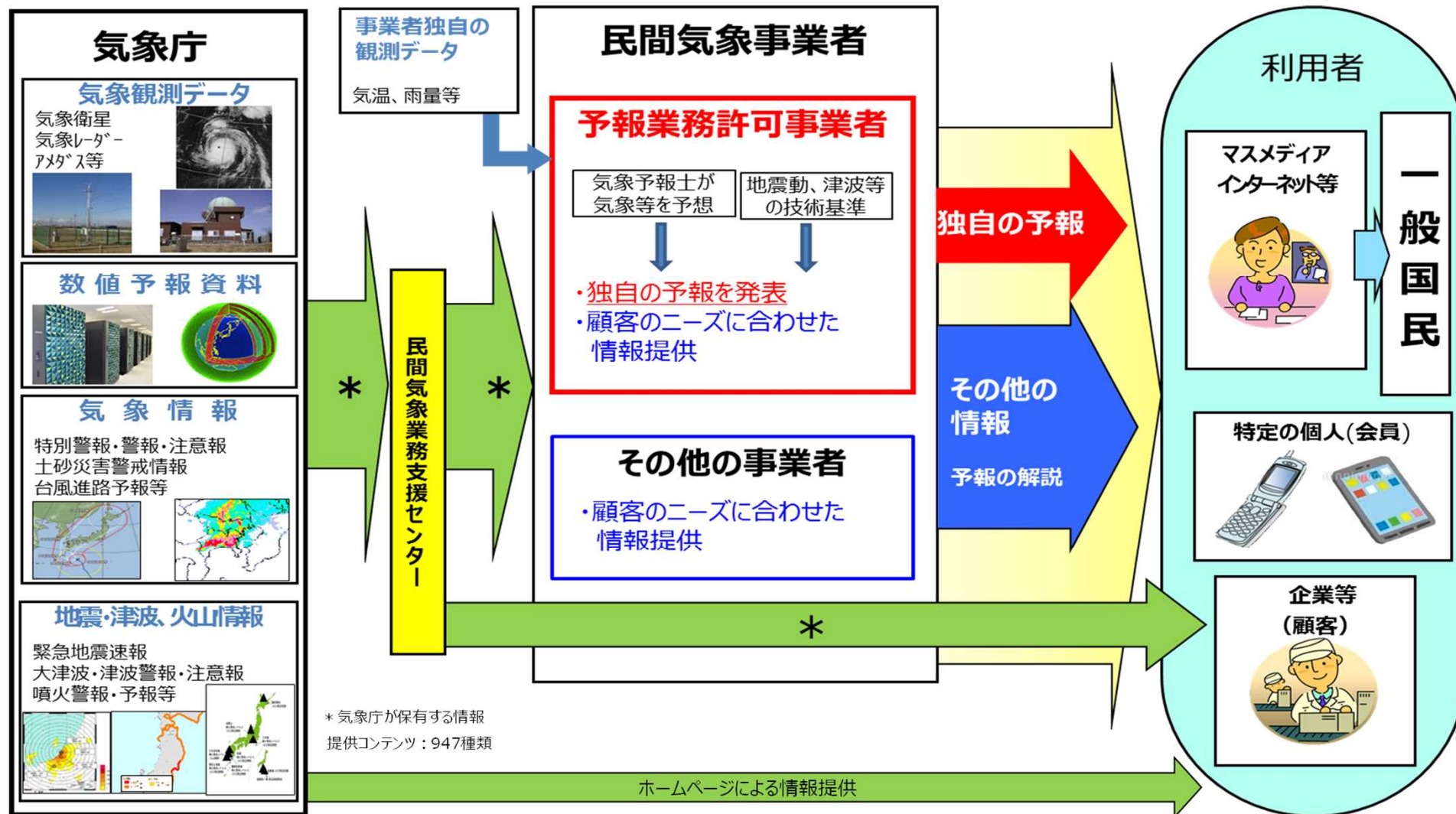
- 直接的な因果関係がないにも関わらず、間接的に相関が強くなる場合があります。これを**疑似相関**と言います。仮説の検証の際には、疑似相関に気をつける必要があります。

講演内容

- 気象データの分析の流れ
- 気象×〇〇の分析例
- 様々な気象データ
- 様々な分析手法
- 気象データ分析に挑戦してみませんか？

内容：気象データを使ってデータ分析してみたい！そんなときに注意すべきポイントとは？ WXBC人材育成WGが企画する「気象データ分析チャレンジ！」の取り組みを例に気象データ活用によるデータ分析の流れを説明するとともに、様々な気象データや様々な分析手法の中からどう使い分けていくのか解説します。みなさんのビジネスに気象データをもっと活用してみませんか？

気象データの入手方法



気象データの種類

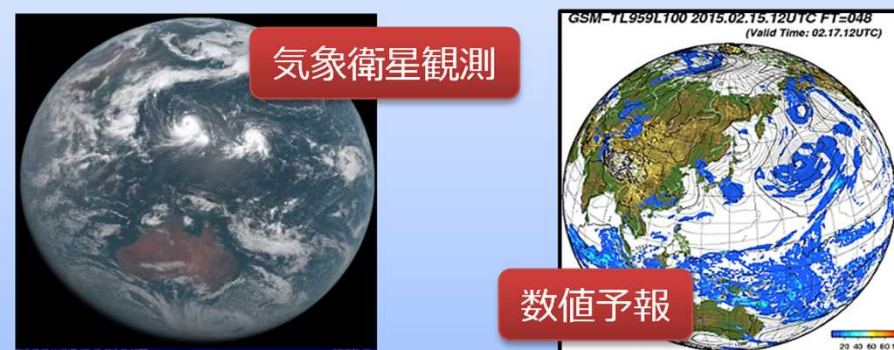
① 全国を網羅する多種多様な気象データ

- アメダス、高層気象観測、天気予報、注意報・警報など、地点・地域の観測・予測データ



② 面的・立体的な広がりを持つ気象データ

- 衛星やレーダー等のメッシュ状の観測データ
- 数値予報等のメッシュ状（3次元）の予測データ



秒・分・時・日・月・年など、様々な時間単位で更新

天気予報、注意報・警報等

- ✓ XML形式等で配信

地点毎データ等

- ✓ BUFR※形式等国際ルールに基づいた形式で配信
- ✓ 過去の気象データをCSV形式で提供

メッシュデータ等

- ✓ GRIB☆形式等国際ルールに基づいた形式で配信

※BUFR : FM94 BUFR 二進形式汎用気象通報式
 ☆GRIB2 : FM92 GRIB 二進形式格子点資料気象通報式(第2版)

ecCodesでデコード可能
 wgrib2やecCodesでデコード可能

気象データの種類

電文データ

文章化された情報を含むデータ（気象警報・注意報等）を、機械判読に適した形式（XML形式）で提供

【気象警報・注意報】

気象特別警報／警報／注意報、土砂災害警戒情報、記録的短時間大雨情報、台風に関する情報、高温注意情報 等



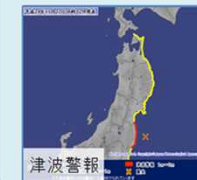
【予報】

今日明日の天気予報、週間天気予報、異常天候早期警戒情報、季節予報（1か月予報、3か月予報、暖・寒候期予報） 等

東京地方	気象情報	降水確率	気温予報
今日25日	北の風 後 北東の風 ぐもり 時々 雨 波 0.5メートル	00-06 100% 06-12 50% 12-18 50% 18-24 50%	日中の最高 23度
明日26日	北東の風 雨 夕方 から ぐもり 波 0.5メートル	00-06 50% 06-12 70% 12-18 50% 18-24 30%	朝の最低 17度 日中の最高 21度
明後日27日	南の風 晴れ 時々 ぐもり 波 0.5メートル		

【地震・津波・火山】

地震情報（震源・震度等）、津波警報／注意報／予報、噴火警報／注意報、噴火速報、降灰予報 等



数値データ

様々な観測や予測による得られた3次元/メッシュデータ等を、国際的ルール（BUFR形式やGRIB形式）で提供

【気象衛星】

ひまわり標準データ、NetCDFデータ、衛星画像（JPEG形式）、カラー画像（PNG形式）、高分解能雲情報 等



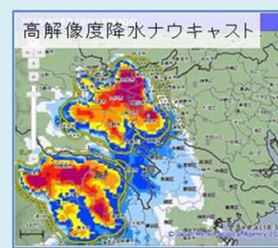
【観測】

アメダス（気温、降水量等）、レーダー（エコー強度等）、雷観測データ、紫外線、潮位実況報 等



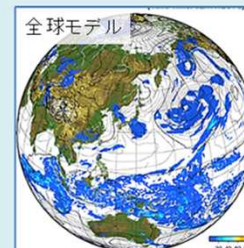
【ナウキャスト】

高解像度降水ナウキャスト、竜巻発生確度ナウキャスト、雷ナウキャスト 等



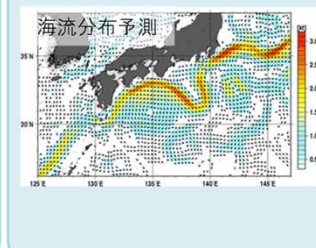
【予測（気象）】

全球モデルGPV、メソモデルGPV、局地モデルGPV、アンサンブルGPV（週間／1か月／3か月予報等）、土砂災害警戒判定メッシュ情報 等



【予測（海洋）】

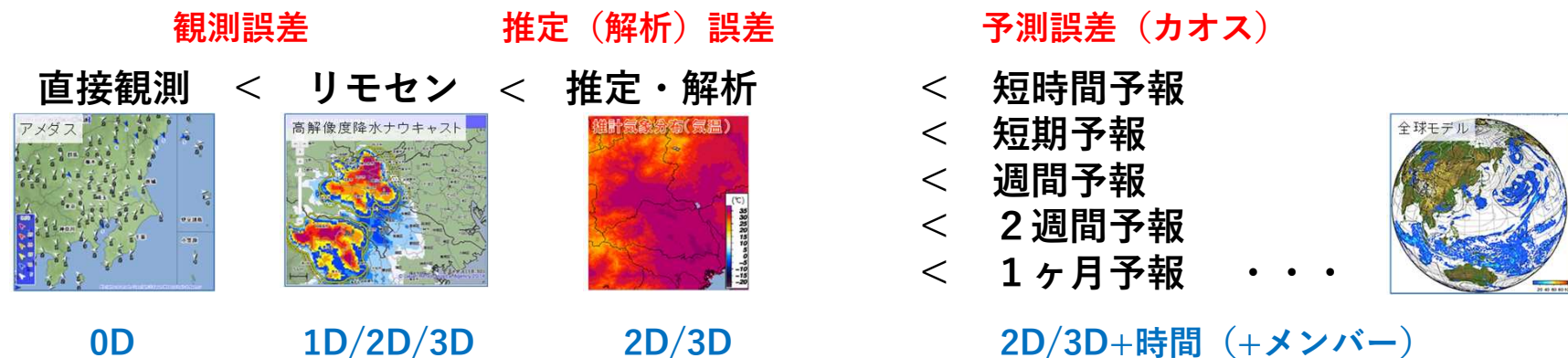
海水温・海流予報GPV、波浪数値予報モデルGPV、地方海上分布予報 等



気象データの有効活用のために

- どんな気象データにも一長一短があります。多種多様な気象データを有効に利用するためには、それぞれのデータが如何にして作成されているのかを把握して、以下のポイントに気をつけながら使い分けてゆく必要があります。

1. 現象は？（気圧？気温？風向／風速？降水？日照時間？天気？）
2. 期間は？（現在？10分先？1時間先？1日先？1週間先？1ヶ月先？）
3. 予報形式は？（量的予報？カテゴリー予報？確率予報？）
4. 予報誤差は？



リードタイムが長くなると、予測誤差が拡大するため確率予報として提供される

気象データを使ってみたい

気象データ、各種情報をカタログとして掲載し、必要な情報を検索し、入手方法を知ることができます。

<https://www.data.jma.go.jp/add/suishin/catalogue/catalogue.html>

メッシュデータ（GRIB2形式）のサンプルデータも入手できます。

https://www.data.jma.go.jp/developer/gpv_sample.html

気象庁情報カタログ

気象庁情報カタログは、気象庁が保有・提供する各種情報(気象情報)のカタログであって、気象情報の利用促進を目的として作成するものです。気象情報を網羅的に記述するとともに、その提供方法についても紹介しています。
現在掲載している内容は概ね令和2年7月時点のものになりますが、可能な範囲で内容を更新しており、実際に提供している気象情報と仕様等が異なる場合がありますので、ご注意ください。

解説

▶ [気象庁情報カタログについて](#)

分野別に表示する



気象



地球環境・気候



海洋



地震・津波



火山



その他

検索する

※ チェックした項目を and 検索します。

提供方法	☐ 気象業務支援センター(オンライン配信) ☐ 気象業務支援センター(オフライン提供) ☐ 気象庁HP ☐ 気象官署等における閲覧
即時提供時のデータ形式	☐ XML ☐ バイナリ ☐ かな漢字 ☐ A/N ☐ カナ ☐ 画像 ☐ FAX
検索	リセット

全ての気象情報を表示する

[全ての気象情報を表示](#)

リンク

▶ [配信資料に関する技術情報](#) < 既存の情報の仕様変更や新たに提供する情報の仕様等の技術的な内容を掲載しています。 >

GPVサンプルデータの一覧

各データ名から、対象領域や解像度、データ形式などの詳細が記載されている「気象庁情報カタログ」をご参照頂けます。

データ名	概要	サンプル
全球数値予報モデルGPV (GSM 全球・日本域)	地球全体の大気を対象に、格子間隔（水平分解能）約20kmとして、未来の気温、風、水蒸気量、日射量等の状態について、スーパーコンピュータを用いて3次元の格子で予測したデータ。132時間先まで（9時、21時（日本時間）初期値のものは264時間先まで）の予測を6時間毎に発表。	サンプル [zip形式：111 MB]
GSMガイドンス (格子形式)	全球数値予報モデルGPV及び観測・解析データから統計手法を用いて作成する、天気、降水量、降水確率などの予報要素を直接示す予測資料。	サンプル [zip形式：344 KB]
GSMガイドンス (最大降水量、降雪量)	全球数値予報モデルGPV及び観測・解析データから統計手法を用いて作成する、最大降水量、降雪量を直接示す予測資料。	サンプル [zip形式：2.52 MB]
GSMガイドンス (視程)	全球数値予報モデルGPV及び観測・解析データから統計手法を用いて作成する、視程を直接示す予測資料。	サンプル [zip形式：292 KB]
メソ数値予報モデルGPV (MSM)	日本とその近海の領域を全球数値予報モデルよりも細かい格子間隔（5km）で、未来の気温、風、水蒸気量、日射量等の状態について、スーパーコンピュータを用いて3次元の格子で予測したデータ。39時間先まで（9時、21時（日本時間）初期値のものに限り51時間先まで）の予測を3時間毎に発表。	サンプル [zip形式：86.3 MB]
MSMガイドンス (格子形式)	メソ数値予報モデルGPV及び観測・解析データから統計手法を用いて作成する、天気、降水量、降水確率などの予報要素を直接示す予測資料。	サンプル [zip形式：1.92 MB]
MSMガイドンス (最大降水量、降雪量)	メソ数値予報モデルGPV及び観測・解析データから統計手法を用いて作成する、最大降水量、降雪量を直接示す予測資料。	サンプル [zip形式：2.17 MB]
MSMガイドンス (視程)	メソ数値予報モデルGPV及び観測・解析データから統計手法を用いて作成する、視程を直接示す予測資料。	サンプル [zip形式：1.70 MB]
局地数値予報モデルGPV (LFM)	メソ数値予報モデルよりさらに細かい格子間隔（2km）で、未来の気温、風、水蒸気量、日射量等の状態について、スーパーコンピュータを用いて3次元の格子で予測したデータ。10時間先までの予測を1時間毎に発表。	サンプル [zip形式：48.1 MB]
	地球全体の大気を対象として、調間単位の気温、風、水蒸気等の状態について、	

気象データを使ってみたい

昨日までのアメダスの気象観測データについて、取得した地点や期間、データの種類を選択して、CSVファイルとしてダウンロードすることができます。

<https://www.data.jma.go.jp/gmd/risk/obsdl/>

検索条件

選択済みのデータ量 0%

100% (上限)

地点を選ぶ

項目を選ぶ

期間を選ぶ

表示オプションを選ぶ

すべての選択済みの地点をクリア

- 一回にリクエストできるデータ量には上限があります(右上棒グラフ参照)。
- アクセス集中時や一回のデータ量が多い場合、繋がらないことやデータ取得に時間がかかることがあります。繋がらない場合は時間を置いて再度お試しください。
- アクセス集中の原因となりますので、自動化ツール等による過度のアクセスは控えたい様お願いいたします。

まず、都道府県を選んでください

宗谷

留萌

上川

網走・北見・紋別

空知

後志

石狩

日高

十勝

根室

釧路

渡島

青森

秋田

岩手

山形

宮城

福島

新潟

富山

石川

福井

滋賀

岐阜

長野

山梨

埼玉

茨城

東京

千葉

神奈川

静岡

愛知

三重

和歌山

奈良

大阪

京都

兵庫

岡山

広島

鳥取

島根

山口

福岡

佐賀

長崎

熊本

大分

愛媛

香川

高知

徳島

鹿児島

宮崎

沖縄

南極

画面に表示 ▶

CSVファイルをダウンロード ▶

選択地点・項目をクリア

選択された地点

観測項目

← 地点を選択してください

選択された項目

← 項目を選択してください

選択された期間(日本標準時)

2021年1月1日から

2021年2月1日までの特別値を表示

選択されたオプション

利用上注意が必要なデータを表示させる

観測環境などの変化以前のデータを表示させる

ダウンロードデータはすべて数値で格納

🔗 ご利用にあたっての注意点

講演内容

- 気象データの分析の流れ
- 気象×〇〇の分析例
- 様々な気象データ
- 様々な分析手法
- 気象データ分析に挑戦してみませんか？

内容：気象データを使ってデータ分析してみたい！そんなときに注意すべきポイントとは？ WXBC人材育成WGが企画する「気象データ分析チャレンジ！」の取り組みを例に気象データ活用によるデータ分析の流れを説明するとともに、様々な気象データや様々な分析手法の中からどう使い分けていくのか解説します。みなさんのビジネスに気象データをもっと活用してみませんか？

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）

✓回帰分析、重回帰分析、決定木分析

✓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

✓ARMA、ARIMA、SARIM、カルマンフィルタ

✓LSTM-RNN

．．．予測的分析

□意思決定分析

✓決定木分析、アソシエーション分析、決定分析

．．．処方的分析

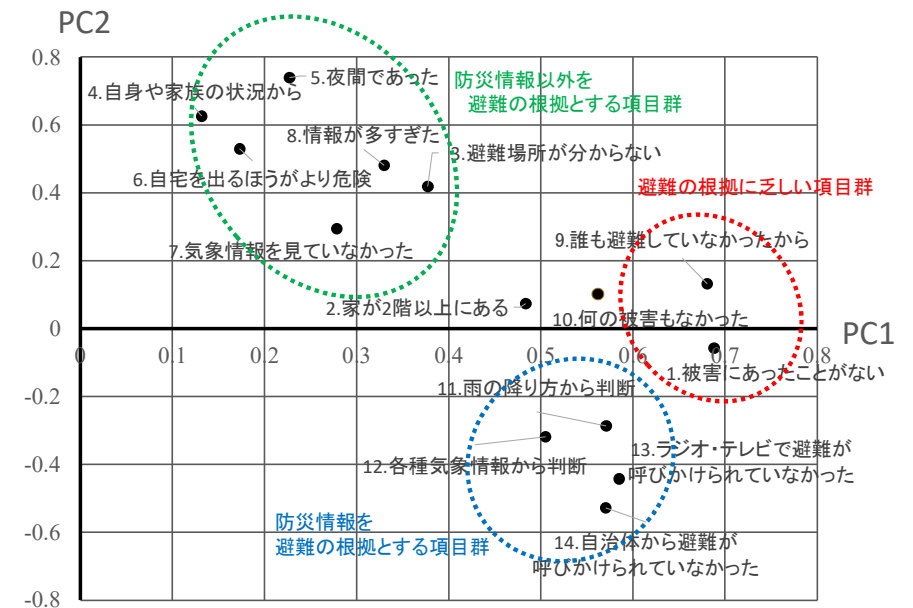
主成分分析

- 主成分分析は、多次元の説明変数からなるデータの次元を圧縮して少ない説明変数でデータ全体の特徴を表す手法です。少ない次元にして説明ができればデータに潜む特徴を理解しやすくなります。

長良川流域住民に「2018年西日本豪雨時に避難しなかった理由」をアンケート調査（222部）の各項目に対して主成分分析を実施

x_1	家が2階以上もしくは高台にあるから
x_2	どこに避難したらよいのか分からなかったから
x_3	自身や家族の状況（高齢等の理由）から避難は簡単ではなかったから
x_4	夜間であり避難は簡単ではなかったから
x_5	自宅を出るほうがより危険だと感じたから
x_6	防災気象情報を見ておらず状況を把握していなかったから
x_7	情報が多すぎて何に従ったらよいかわからなかったから
x_8	過去に一度も被害にあったことはなかったから
x_9	家族・近所の人・知人が誰も避難していなかったから
x_{10}	停電・断水・損壊など自宅に何の被害もなかったから
x_{11}	雨の降り方や河川や斜面の様子から特に身の危険を感じなかったから
x_{12}	各種の気象情報や河川情報から危険でないと判断したから
x_{13}	テレビやラジオなどで避難が呼びかけられていなかったから
x_{14}	自治体から避難が呼びかけられていなかったから

第1主成分（PC1）と第2主成分（PC2）の各変数の因子負荷量のプロット（バイプロット）。多数のパラメータを2個の主成分に集約。



クラスター分析

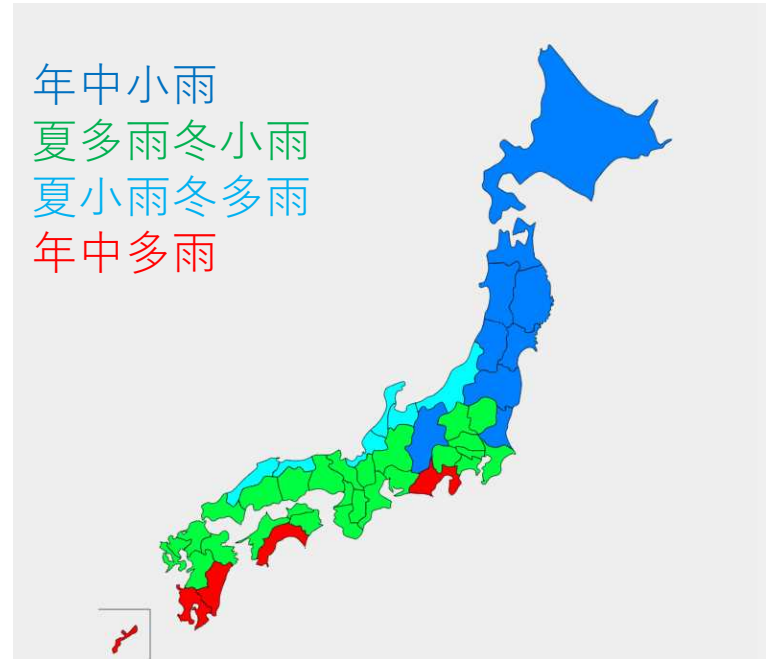
- クラスター分析とは、多数のサンプルからなるデータをグループ（クラスター）に分類するための手法です。全体のユークリッド距離（類似性）が最小となるグループ化を行う。グループ毎に説明変数の統計値を比較することで、比較しやすくなります。

気象庁より気温と降水量の平年値を入手して、県庁所在地毎に4つのクラスターに分類（K-means法）

	A	B	C	D	E	F	G	H	I
1		気温DJF	気温MAM	気温JJA	気温SON	降水量DJF	降水量MAM	降水量JJA	降水量SON
2	那覇	17.6	21.5	28.1	25	327.7	558.7	629.1	523.6
3	長崎	8	15.2	25.9	19.6	207.6	462.6	824.5	360.2
4	佐賀	6.5	14.8	26	18.5	179	483	874.5	330.8
5	熊本	6.8	15.5	26.4	19	195.1	479.3	979.2	330.4
6	鹿児島	9.6	16.7	26.9	21.1	259.4	605.5	994.1	405.2
7	宮崎	8.5	16	25.8	19.4	212.4	633.8	1028.8	631.3
8	大分	7.2	14.3	25.4	18.7	143.8	391.6	698.5	409.5
9	福岡	7.6	15	26.1	19.1	196	371.6	704.7	336.8
10	山口	5.2	13.6	25.1	17.3	213.9	526.4	801.3	343.3
11	広島	6.2	14.4	26.1	18.4	150.8	443.2	616.4	325.6
12	松江	5.3	12.7	24.5	17	404.8	376.6	555.9	448
13	鳥取	5	12.7	24.8	17	552.2	381.1	469.6	507.6
14	岡山	5.9	14.2	26.3	18.3	114.6	303.9	419.8	266.7
15	松山	6.9	14.3	25.8	18.8	161.9	351.7	504.7	295.1
16	高松	6.4	14.2	26.1	18.5	122.6	266.6	380.4	312.2
17	徳島	7	14.5	25.7	19	133.8	351.1	512.5	453.4
18	高知	7.4	15.4	25.7	19.3	218.6	726.4	957.2	640.7
19	神戸	6.9	14.5	26.1	19.4	135.4	349.8	424.6	306.3
20	京都	5.6	13.9	26	18	164.9	389.8	566.5	368.4
21	大阪	6.9	14.7	26.6	19.2	149.5	353.4	432.4	342.4
22	和歌山	6.9	14.6	26	19	152.5	346.8	419.4	395.9
23	奈良	4.8	13	24.9	16.9	158.7	344.4	465.6	345.8
24	福井	4.1	12.4	24.8	16.8	733.4	430.3	527.5	552.4
25	彦根	4.6	12.1	24.7	17.2	299.6	384.7	517.2	368.9

各都道府県のクラスター分析による分類結果。よく知られている気候区分通りに分類できた。

- A) 寒冷 年中小雨
- B) 温暖 夏多雨冬小雨
- C) 温暖 夏小雨冬多雨
- D) 高温 年中多雨

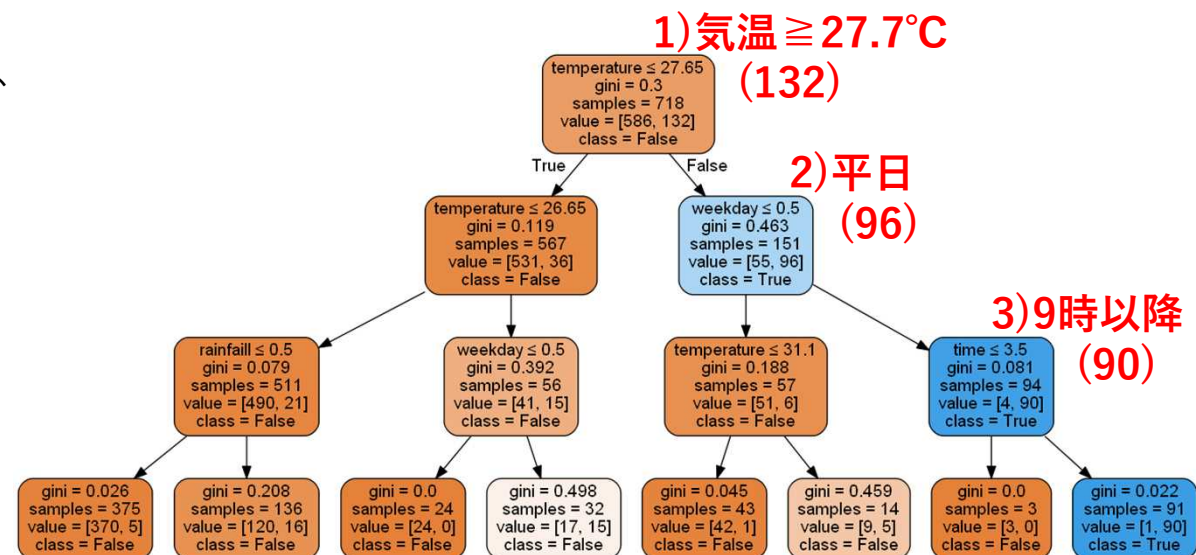


決定木分析

- 決定木分析とは、多数のサンプルデータの中から目的変数に対してきれい（純度の高い）に分類できる決定木（樹木状のモデル）を作って分類や回帰を行う手法。顧客情報やアンケート結果などから目的変数に影響する説明変数を見つけることができる。

2016年7月（夏）の東京電力の電力消費量4000万kWを超える時間について、曜日帯、時間帯、気温、降水量により決定木（3層まで）を作成（CART法）

	A	B	C	D	E	F	G	H
1	月	日	時間	電力消費量4000万kW以上	平日	時間帯	気温	降水有無
2	7	1	1	FALSE	1	1	22.8	0
3	7	1	2	FALSE	1	1	22.6	0
4	7	1	3	FALSE	1	1	22.3	0
5	7	1	4	FALSE	1	2	22.3	0
6	7	1	5	FALSE	1	2	22.7	0
7	7	1	6	FALSE	1	2	22.7	0
8	7	1	7	FALSE	1	3	23.2	0
9	7	1	8	FALSE	1	3	24.2	0
10	7	1	9	FALSE	1	3	24.8	0
11	7	1	10	FALSE	1	4	25.6	0
12	7	1	11	FALSE	1	4	26.1	0
13	7	1	12	TRUE	1	4	27.2	0
14	7	1	13	TRUE	1	5	28.6	0
15	7	1	14	TRUE	1	5	29.6	0
16	7	1	15	TRUE	1	5	29.4	0
17	7	1	16	TRUE	1	6	29.8	0
18	7	1	17	TRUE	1	6	29.3	0



夏は、1) 気温が 27.7°C 以上、2) 平日、3) 午前9時以降、で電力消費量は4000万kW以上になる傾向がある。**132事例中の90事例。**

ニューラルネットワーク

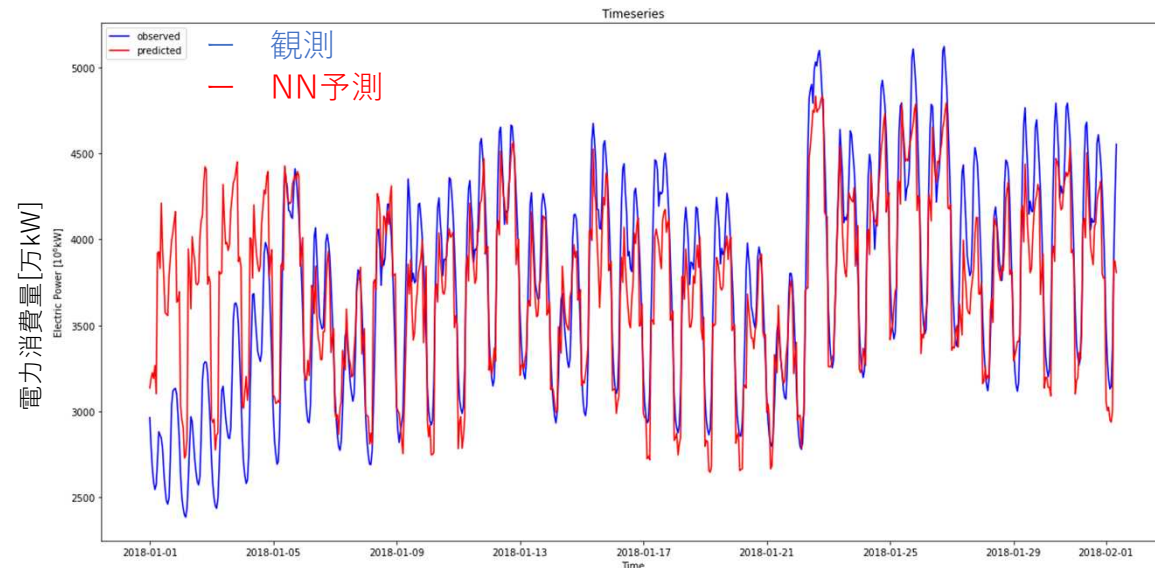
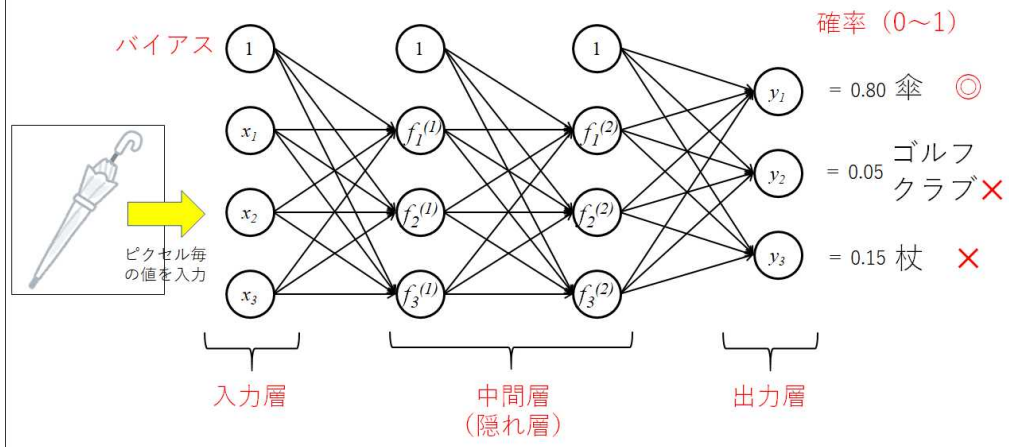
アメダス気象データ分析チャレンジ！（Python版）より

- ニューラルネットワークとは、神経細胞を模した多数のニューロンを組み合わせたネットワークを作成し、入力層と出力層の関係からニューロン間の重みを最適化することで得られる機械学習モデルの一種。活性化関数やニューロンの多層化により高精度な予測が可能。

ニューラルネットワークの構造。出力層の構造を変化させることで分類問題や回帰問題に適用できる。

ニューラルネットワークにより、2018年1月の電力消費量の変化を予測。入力層は、気温と曜日帯と時間帯からなる。**決定係数：0.86**

ニューラルネットワークとは？

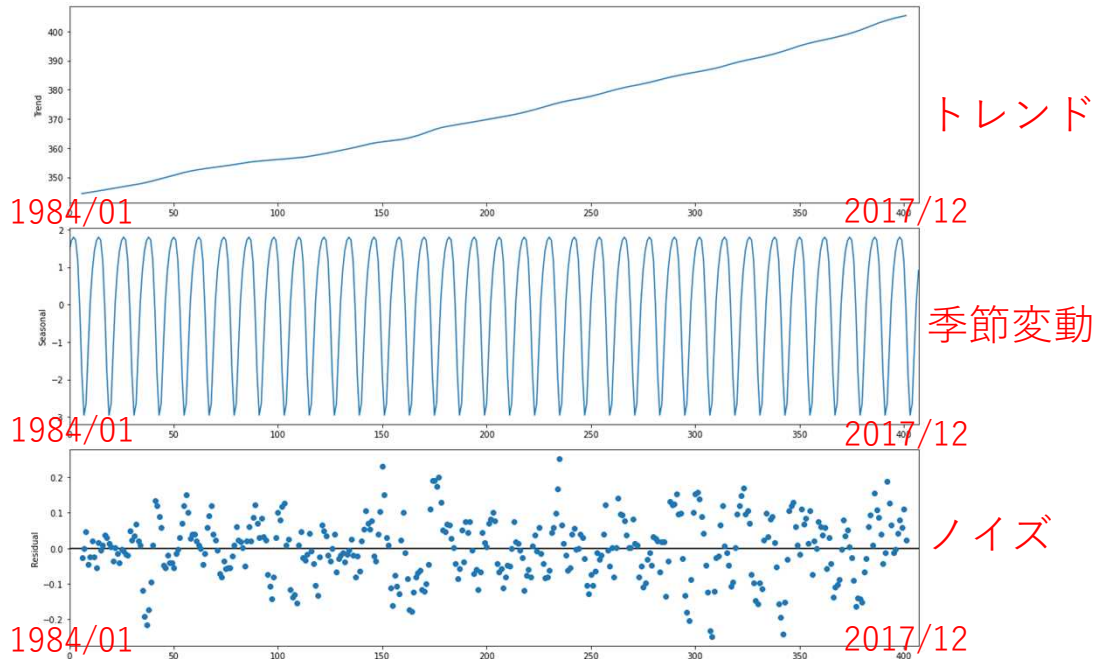


2018年1月

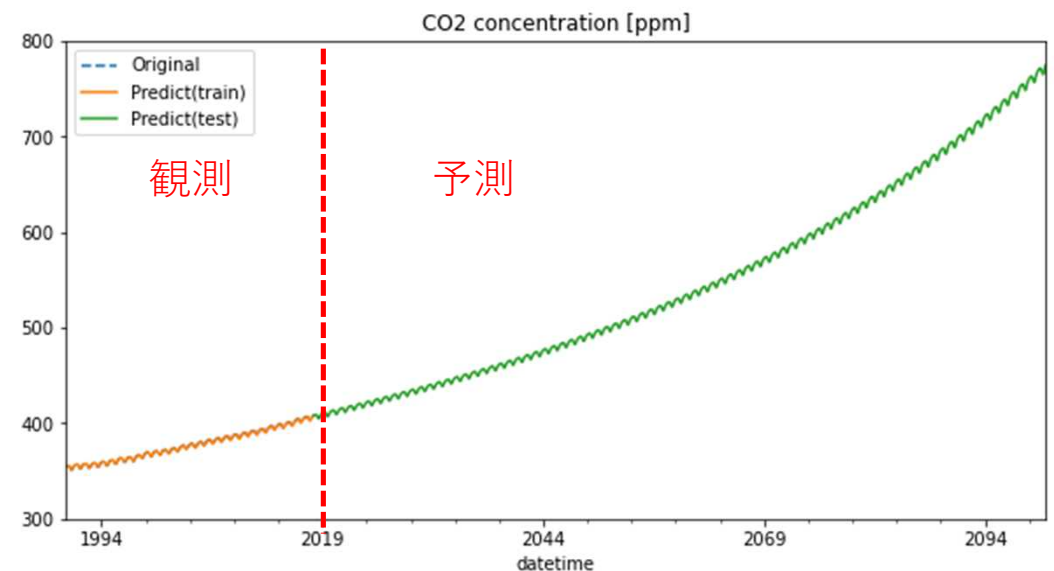
ARMAモデル

- ARMAモデル（自己回帰移動平均モデル）は、自己回帰モデルAR（非定常過程を表現）と移動平均モデルMA（定常過程を表現）を組み合わせた時系列モデルです。時間変動の分析に用いる。トレンドを含めたARIMAモデル、季節変動を含めたSARIMAモデルがある。

全世界平均の大気中CO2濃度（気象庁）の時系列から、ARMAモデルによる分解された成分。



SARIMAモデルにより、2100年末までのCO2濃度の予測結果。このペースで排出すると、2000年時の倍近いCO2濃度になる。

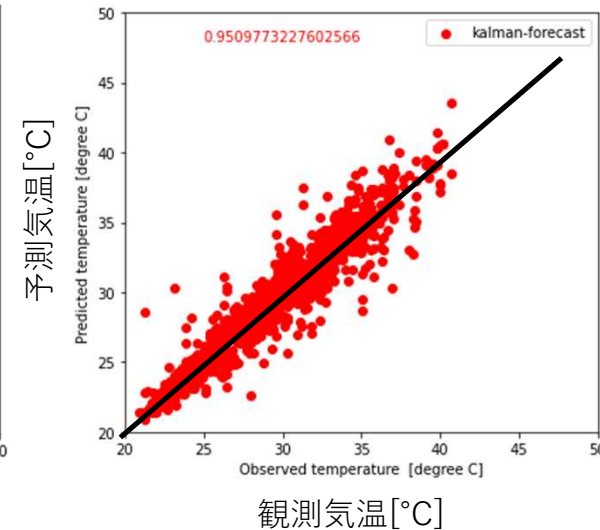
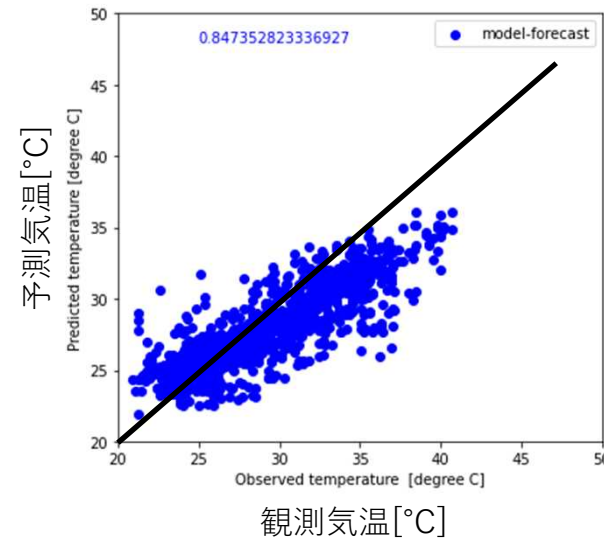
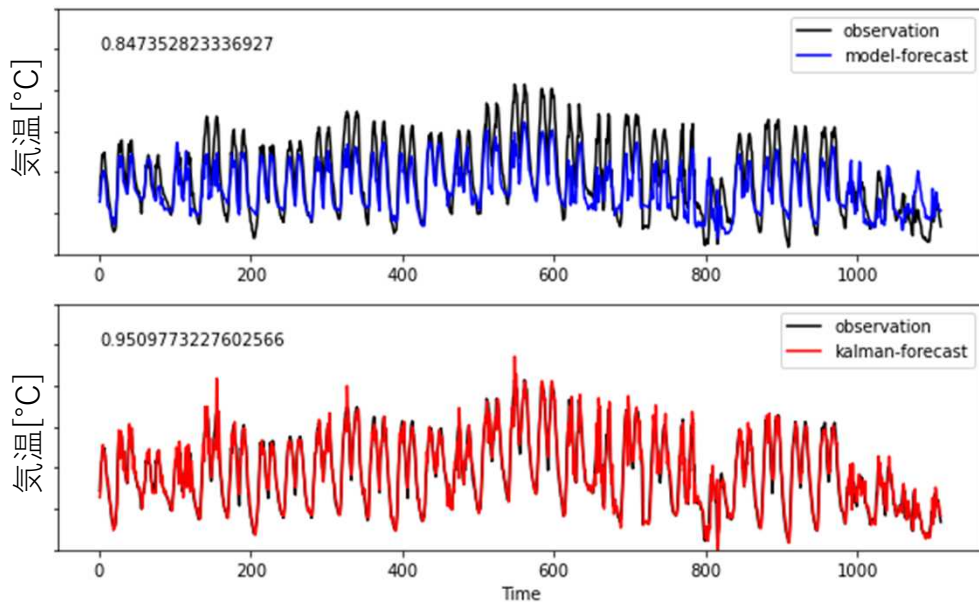


カルマンフィルタ

- カルマンフィルタとは、定常・非定常問わず、システムモデルと観測モデルに基づき、ノイズ混じりのデータから、1期先予測とフィルタを繰り返して逐次的に補正する手法。気象庁の天気予報ガイダンス（モデルの予測を過去の観測によって補正する）でも利用されている。

ある地点の気象モデル（36時間）による気温の予測結果（青色の線）に対して、カルマンフィルタにより直前までに観測された気温で補正（赤色の線）

モデルによる未来の予測データと過去の観測データを組み合わせることで、モデル x と誤差 y の線形回帰式（ $y = ax + b$ ）により修正（高精度化）できる。



決定分析

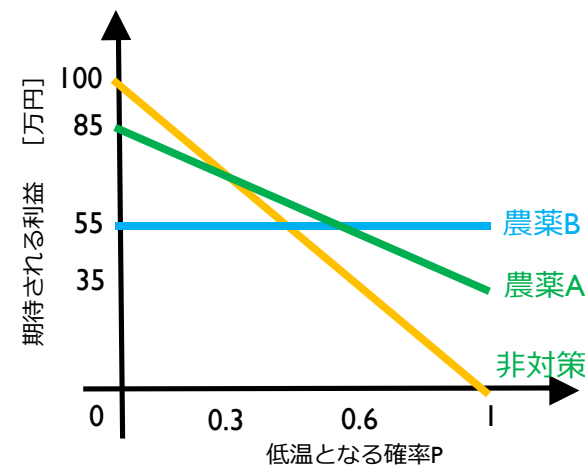
- 決定分析とは、天気予報のように不確実な情報（確率予報）から期待される利益（損失の軽減）を最大にする最適な対策を探るための分析手法。
MINIMAX（リスク回避型）、MINIMIN（リスク放置型）、MINIMAX・REGRET（後悔回避型）など意思決定者が何を重視するかで決まる。

低温になると病虫害の発生が懸念される農作物がある
とします。病虫害がなければ**100万円の収穫**が期待でき、
ひとたび病虫害が発生すれば**100万円の損害**が発生
するとします。以下のような**2種類の農薬Aと農薬B**
があるとき、低温となる確率Pが何%のとき、どう対
応すべきでしょうか？

農薬ごとのコストと効能

対策	コスト	効能
農薬A	15 万円	農作物の被害を50万円に軽減できる
農薬B	45 万円	農作物の被害を完全に防止できる
非対策	0 万円	被害は100万円となる

各対策毎に期待利益（縦軸）を低温確率P（横軸）の関数で表し、最も期待利益の高くなる対策を選択する。



低温の確率毎の最適な対策

低温の確率	取るべき対策
10%	非対策
20%	非対策
30%	農薬Aを散布
40%	農薬Aを散布
50%	農薬Aを散布
60%	農薬Bを散布
70%	農薬Bを散布
80%	農薬Bを散布

講演内容

- 気象データの分析の流れ
- 気象×〇〇の分析例
- 様々な気象データ
- 様々な分析手法
- 気象データ分析に挑戦してみませんか？

内容：気象データを使ってデータ分析してみたい！そんなときに注意すべきポイントとは？ WXBC人材育成WGが企画する「気象データ分析チャレンジ！」の取り組みを例に気象データ活用によるデータ分析の流れを説明するとともに、様々な気象データや様々な分析手法の中からどう使い分けていくのか解説します。みなさんのビジネスに気象データをもっと活用してみませんか？

気象データ分析に挑戦してみませんか？

- 気象データ分析には、「気象データ理解力」「IT活用
力」「ビジネス課題解決力」が必要。皆さんも気象データ
アナリストを目指してみませんか？
- データ分析とは、仮説と検証の繰り返しによりデータに潜
む本質に迫ること。データを入手し、データを加工・整理
し、データを可視化し、様々な手法で分析します。
- 様々な気象データと様々な分析手法があります。皆さんの
ビジネスの目的に応じて、これらを賢く使い分けてゆく必
要があります。
- WXBC人材育成WGでは、皆さんの気象データ分析のため
の学びの場を提供してゆきます。乞うご期待。

気象データ分析に挑戦してみませんか？

動画で気軽に復習したい方へ：

●アメダス気象データ分析チャレンジ！ e-learning

URL : <https://www.wxbc.jp/weather-challenge/index.html>



気象データ分析に挑戦してみませんか？

オンラインで一緒に学びたい方へ：

●アメダス気象データ分析チャレンジ入門

日時：令和3年9月6日（月）午後（3時間程度の予定）

研修内容：Excelにより気象データと電力データをかけ合わせた分析を行います

講師：吉野 純（岐阜大学工学部附属応用気象研究センター）
気象データ×IT勉強会有志

対象者：気象データ分析に関心のある方（WXBC会員・非会員を問わず）

開催形態：オンラインのハンズオン研修

募集人数：80名程度（予定）

皆様のご参加をお待ちしております！

ご清聴ありがとうございました

吉野 純

岐阜大学工学部附属応用気象研究センター

E-mail: jyoshino@gifu-u.ac.jp

URL: <http://www1.gifu-u.ac.jp/~amet>

