

気象データ分析チャレンジ！

データ分析の基礎と個人ワーク

Copyright 2021 気象ビジネス推進コンソーシアム、岐阜大学 吉野純

(C) 2021 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。

Contents

- データ分析の手順
- 時系列図でデータを可視化する
- 散布図と相関係数で検証する

Contents

- データ分析の手順
- 時系列図でデータを可視化する
- 散布図と相関係数で検証する

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、何らかの活動に役立てる目的があり、その目的を達成するためにデータ分析をする。まずは、目的を言語化してみよう。

□(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)スキー場の最適な人員配置計画を立てたい

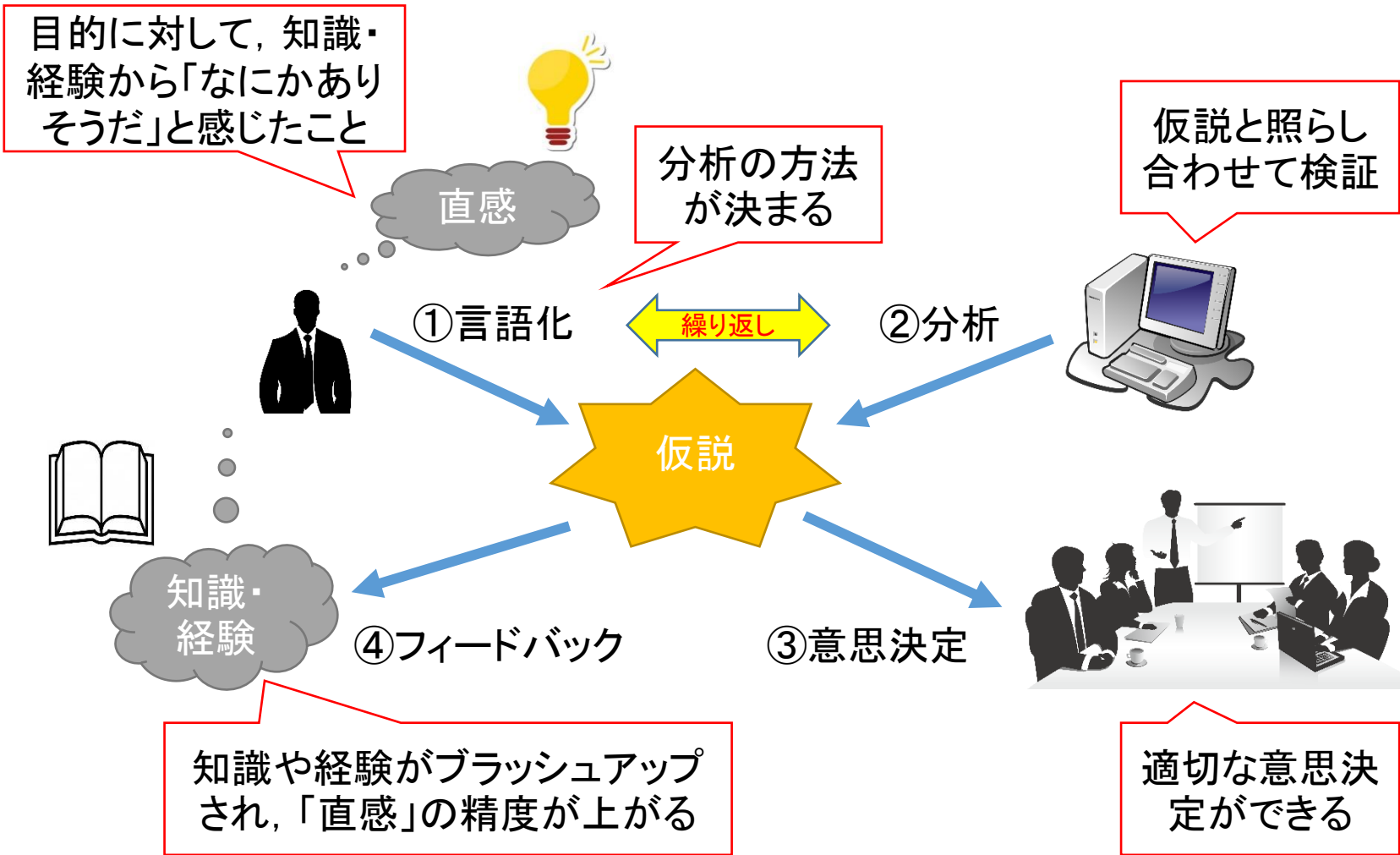
➤(仮説)降雪量や気温と来場者数との間を比較して特徴を見つける

□(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)地域毎の気温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

仮説が正しいかを検証する必要がある！

仮説と検証の繰り返し



仮説と分析の繰り返しにより本質の理解に近づく

エクセルによるデータ分析

□データの要約

大量のデータから、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。

□データの検証

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！

データの要約

- 生データのままでは分析に使えないことが多い。
そこで、必要な項目だけを取り出したり、顧客や商品毎に並べ替えたり、金額を集計したりして、分析に適した形式に加工する。



ファイル ホーム **挿入** ページレイアウト 数式 データ

ピボットテーブル おすすめのピボットテーブル テーブル 画像 オンライン画像 図形 SmartArt スクリーンショット

A3 : X Y fx No.

	A	B	C	D	E
1	新商品モニターアンケートデータ				
2					
3	No.	年齢	性別	味の評価	ロゴ案A~C
4	1	24	F	4 C	
5	2	30	M	3 A	
6	3	15	F	5 A	
7	4	28	M	3 B	
8	5	18	M	5 C	
9	6	24	F	5 C	
10	7	26	F	4 C	
11	8	19	M	5 A	
12	9	27	F	2 C	
13	10	30	M	5 B	

ピボット
テーブル
➡
(クロス集計)

性別毎の味の評価の平均

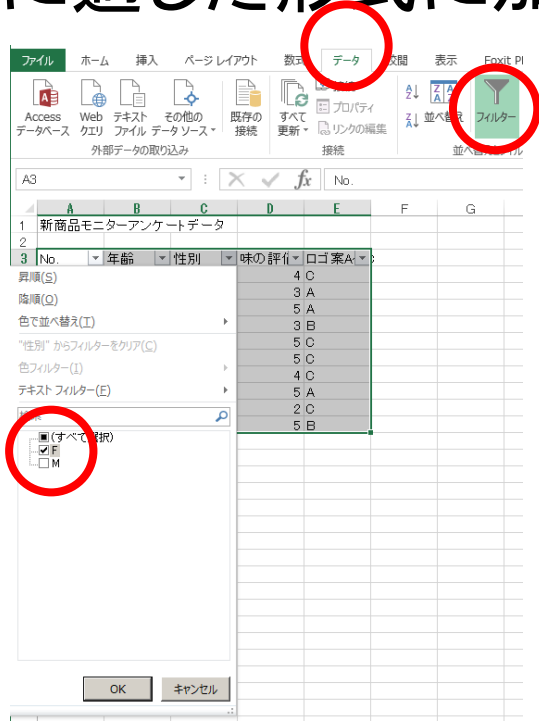
	A	B	C	D
1				
2				
3	行ラベル	平均 / 味の評価		
4	F		4.0	
5	M		4.2	
6	総計		4.1	
7				

性別毎のロゴ案の総計

	A	B	C	D	E
1					
2					
3	個数 / ロゴ案A~C (列ラベル)				
4	行ラベル	A	B	C	総計
5	F	1	0	4	5
6	M	2	2	1	5
7	総計	3	2	5	10

データの要約

- 生データのままでは分析に使えないことが多い。
そこで、必要な項目だけを取り出したり、顧客や商品毎に並べ替えたり、金額を集計したりして、分析に適した形式に加工する。



フィルター
➡

女性のデータだけ抽出できる

	A	B	C	D	E
1	新商品モニターアンケートデータ				
2					
3	No.	年齢	性別	味の評価	ロゴ案A
4	1	24	F	4 C	
6	3	15	F	5 A	
9	6	24	F	5 C	
10	7	26	F	4 C	
12	9	27	F	2 C	

データの要約

- 生データのままでは分析に使えないことが多い。
そこで、必要な項目だけを取り出したり、顧客や商品毎に並べ替えたり、金額を集計したりして、分析に適した形式に加工する。

VLOOKUP関数

H4

=VLOOKUP(G4,\$A\$4:\$E\$13,2,TRUE)

	A	B	C	D	E	F	G	H
1	新商品モニターアンケートデータ							
2								
3	No.	年齢	性別	味の評価	ロゴ案A~C		No.	年齢
4	1	24	F	4	C		2	30
5	2	30	M	3	A		5	18
6	3	15	F	5	A		7	26
7	4	28	M	3	B		10	30
8	5	18	M	5	C			
9	6	24	F	5	C			
10	7	26	F	4	C			
11	8	19	M	5	A			
12	9	27	F	2	C			
13	10	30	M	5	B			

左のデータベースから
一部の人の情報を抽出できる

または、セルの中で関数とデータの配列を直接打ち込む。

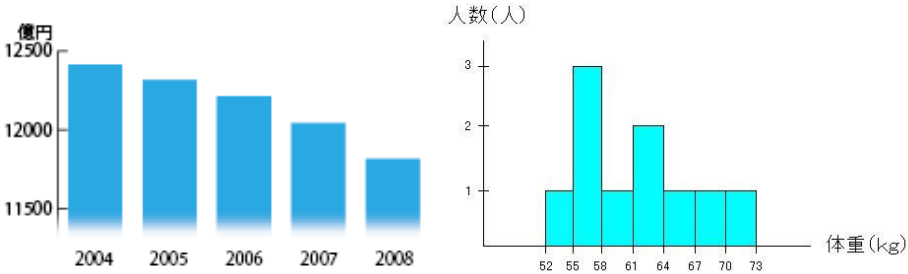
=VLOOKUP(G4 , A3:E13 , 2 , FALSE)

検索値 範囲 抽出列 完全一致

データの可視化

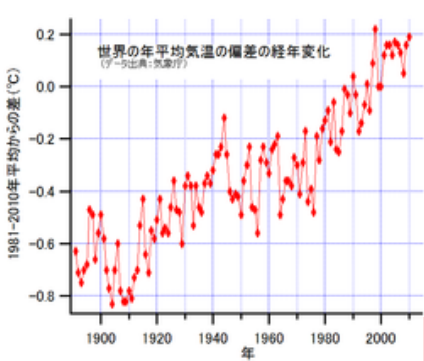
- データを適切な形にグラフ化すると、その特徴がみえてくる。数値だけでは分かりにくい例外的な特徴を抽出することもできる。分析の目的に合ったグラフを選ぶよう注意する。

目的	グラフの種類
数や量の大きさを比較する	棒グラフ
	積み上げ棒グラフ
分布の全体像を見る	ヒストグラム
値の推移を知る	折れ線グラフ
	積み上げ面グラフ



棒グラフ

ヒストグラム

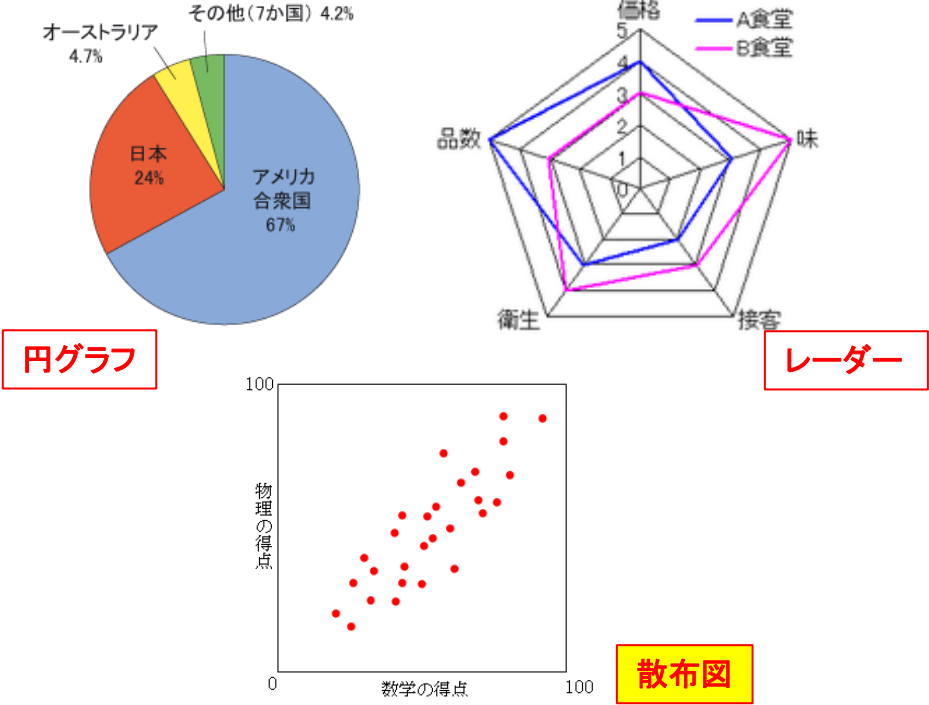


折れ線グラフ

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。

目的	グラフの種類
構成比を見る	円グラフ
	パレート図
2つの変数の関係を見る	散布図
いくつかの項目を元にプロフィールを作る	レーダーチャート



データの検証

- ・ 集団の中から全体像や特徴を見きわめるために、さまざまな統計関数を利用する。

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

データの検証

- 集団の中から全体像や特徴を見きわめるために、さまざまな統計関数を利用する。

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.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

アドインを利用する

- エクセルには標準で用意されている関数の他にも、アドインと呼ばれる拡張機能が用意されている。「分析ツール」にはデータ分析に役立つ様々な機能が含まれている。
- 基本統計量のほか、平均値の差の検定(t検定)、分散の差の検定(F検定)、相関係数、回帰分析、分散分析などができる。
- 「分析ツール」の有効化の方法: 「ファイル」をクリック → 「オプション」をクリック → 「アドイン」をクリック → 「設定」をクリック → 「分析ツール」にチェック → 「OK」をクリック

Contents

- データ分析の手順
- 時系列図でデータを可視化する
- 散布図と相関係数で検証する

データ分析してみよう

例えば,

	A	B	C	D	E
1	気温と商品Aの販売数の関係				
2	日付	気温	販売数		
3	4月1日	12.6	23		
4	4月2日	11.6	27		
5	4月3日	12.3	22		
6	4月4日	14.7	21		
7	4月5日	12.0	32		
8	4月6日	14.7	27		
9	4月7日	13.5	27		
10	4月8日	15.2	24		
11	4月9日	18.6	45		
12	4月10日	17.7	36		
13	4月11日	11.7	32		
14	4月12日	10.5	26		
15	4月13日	15.5	24		

(注意！) 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多いです。

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

(仮説) 商品Aの販売数と気温との間の関係を探ってみよう

エクセルで時系列図

	A	B	C	D	E
1	気温と商品Aの販売数の関係				
2	日付	気温	販売数		
3	4月1日	12.6	23		
4	4月2日	11.6	27		
5	4月3日	12.3	22		
6	4月4日	14.7	21		
7	4月5日	12.0	32		
8	4月6日	14.7	27		
9	4月7日	13.5	27		
10	4月8日	15.2	24		
11	4月9日	18.6	45		
12	4月10日	17.7	36		
13	4月11日	11.7	32		
14	4月12日	10.5	26		
15	4月13日	15.5	24		

A列とB列を選択する

	A	B	C
1	気温と商品Aの販売数の関係		
2	日付	気温	販売数
3	4月1日	12.6	23
4	4月2日	11.6	27
5	4月3日	12.3	22
6	4月4日	14.7	21
7	4月5日	12.0	32
8	4月6日	14.7	27
9	4月7日	13.5	27
10	4月8日	15.2	24
11	4月9日	18.6	45
12	4月10日	17.7	36
13	4月11日	11.7	32
14	4月12日	10.5	26
15	4月13日	15.5	24

A列とC列を選択する (Ctrlを押しながら)

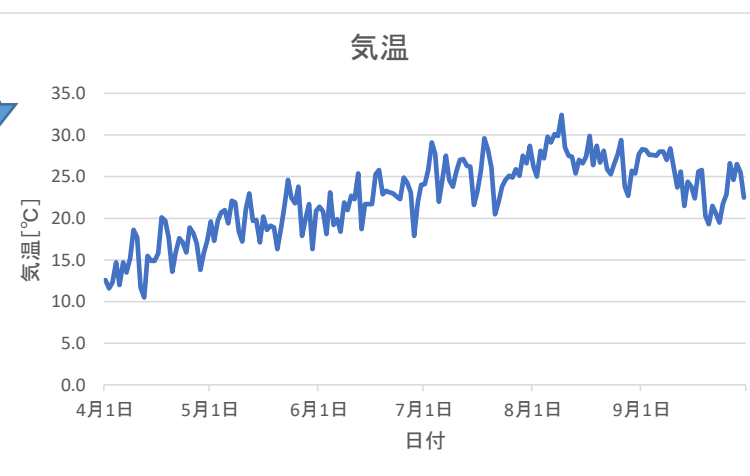
折れ線グラフの挿入

この種類のグラフは、一定の時間 (年数、月数、日数) や項目にわたるデータの傾向を表すのに使います。

別の種類の折れ線グラフを表示するには、矢印をクリックします。ドキュメント内でプレビューを表示するには、ポインターをアイコンの上に置きます。

ツールバーの中から折れ線グラフをクリック

気温



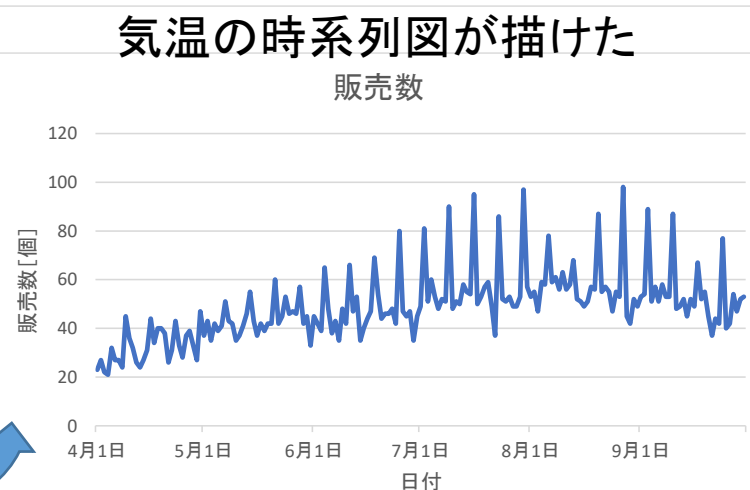
気温 [°C]

4月1日 5月1日 6月1日 7月1日 8月1日 9月1日

日付

気温の時系列図が描けた

販売数



販売数 [個]

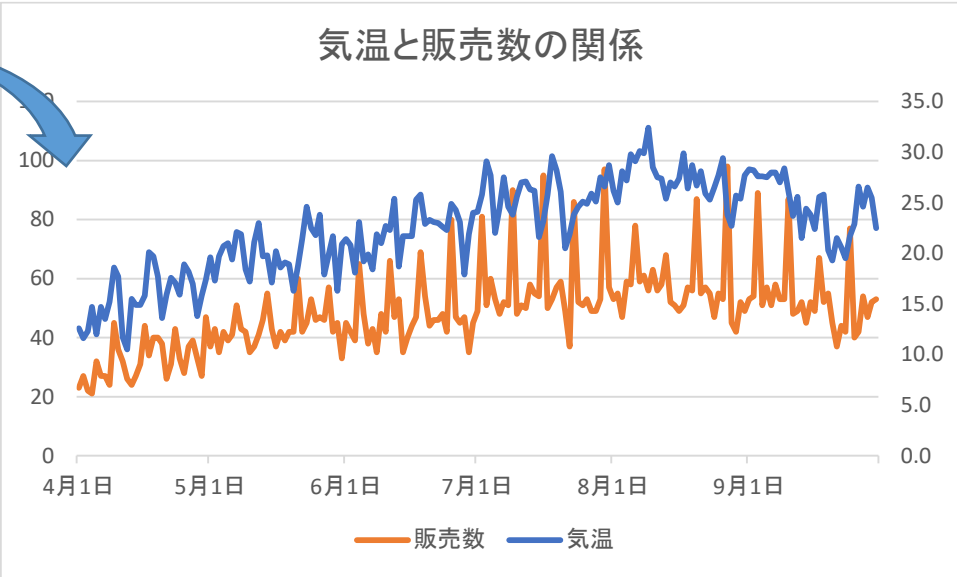
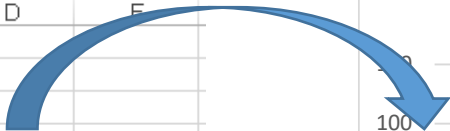
4月1日 5月1日 6月1日 7月1日 8月1日 9月1日

日付

販売数の時系列図が描けた

エクセルで時系列図

	A	B	C	D	E
1	気温と商品Aの販売数の関係				
2	日付	気温	販売数		
3	4月1日	12.6	23		
4	4月2日	11.6	27		
5	4月3日	12.3	22		
6	4月4日	14.7	21		
7	4月5日	12.0	32		
8	4月6日	14.7	27		
9	4月7日	13.5	27		
10	4月8日	15.2	24		
11	4月9日	18.6	45		
12	4月10日	17.7	36		
13	4月11日	11.7	32		
14	4月12日	10.5	26		
15	4月13日	15.5	24		



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:

- 気温の上昇に応じて販売数が増えている。
- 気温の低い春先に比べて夏は倍近い販売数がある。
- 週に1回、異常によく売れている日がある。日曜特売のせい？

データ分析演習1 ～やってみよう！

- まずは、必要な気象データとオープンデータをダウンロードしてみましよう。今回は、東京電力の「でんき予報」のデータを利用します。

東京電力でんき予報(東京電力パワーグリッドが提供)

<http://www.tepco.co.jp/forecast/index-j.html>

- 2016年のデータ(4月1日0時～12月31日23時)をダウンロードしてみよう。電力使用実績データから9ヶ月間の時系列図を作ってみよう。

データ分析演習1 ～やってみよう！～

- 気象データをダウンロードします。電力使用実績データと「同じ期間」の「東京」の「時別値」の「気温」をダウンロードしよう。9ヶ月間の時系列図を作ってみよう。

気象庁 過去の気象データ・ダウンロード

<http://www.data.jma.go.jp/gmd/risk/obsdl/index.php>

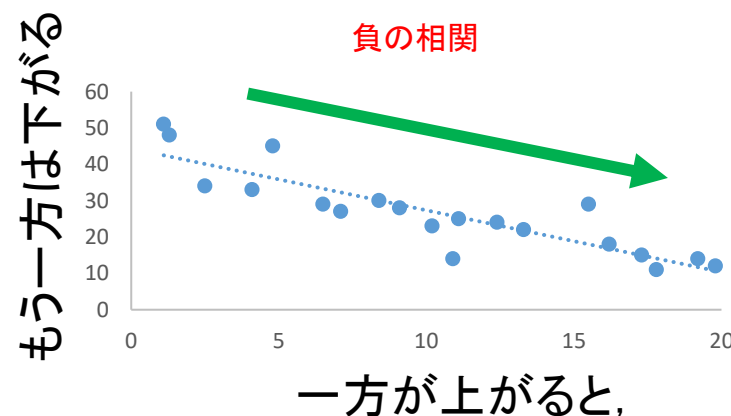
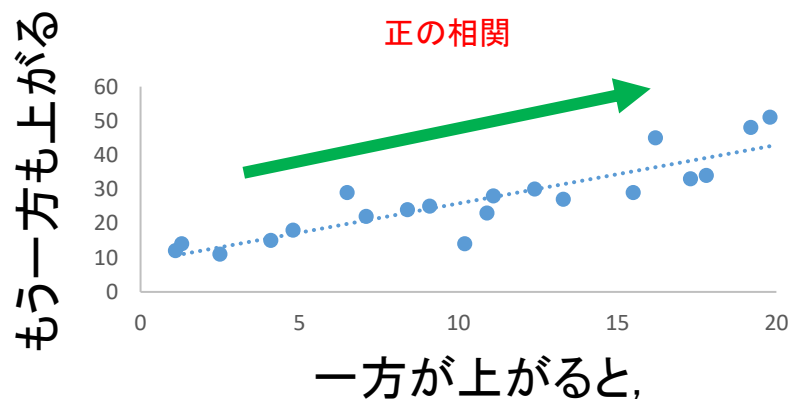
- 気温と電力使用実績を重ねて時系列図で表してみよう。そこからぼんやりと読み取れることを整理してみよう。

Contents

- データ分析の手順
- 時系列図でデータを可視化する
- 散布図と相関係数で検証する

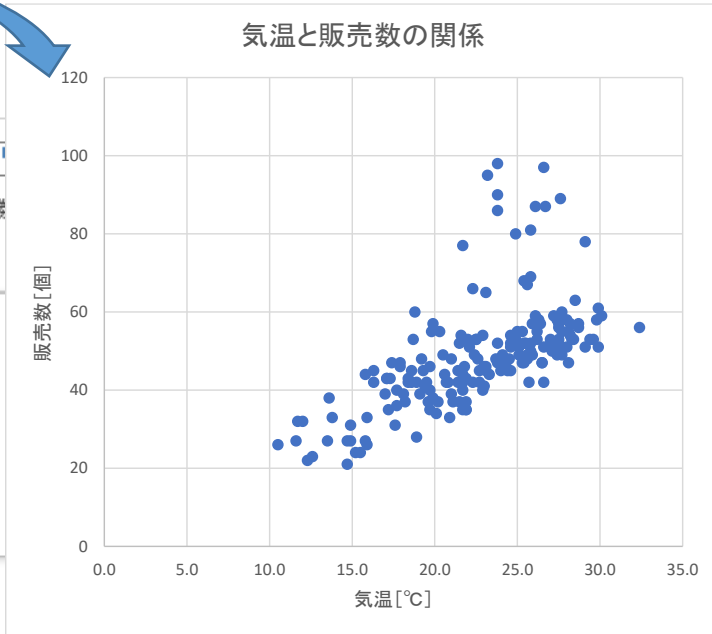
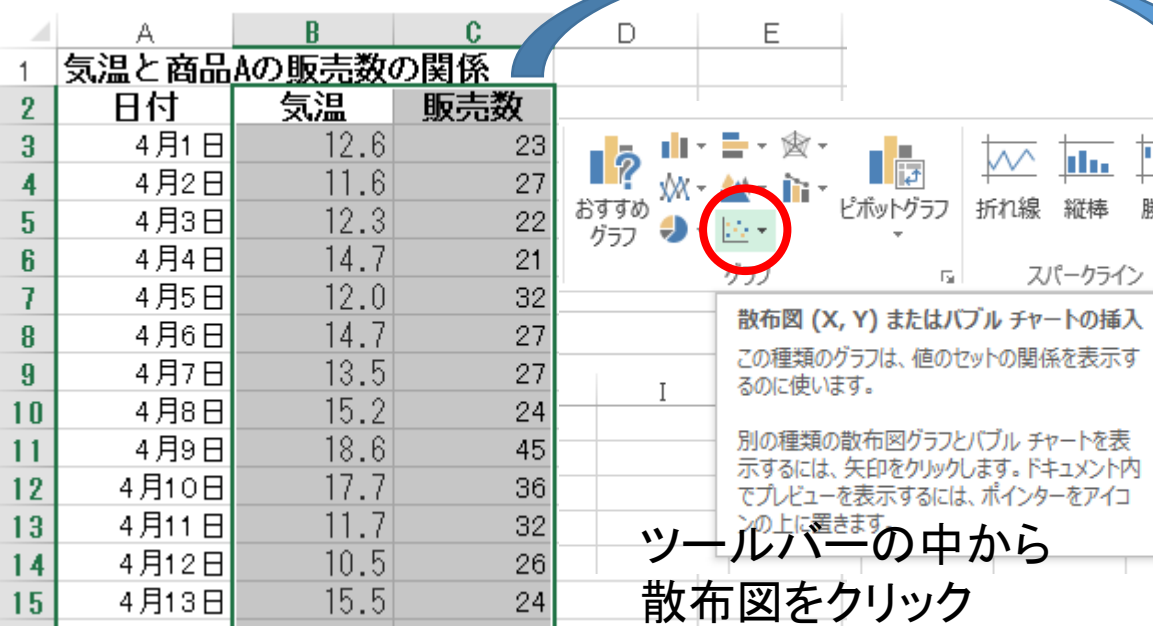
相関関係

- 時系列図から、気温と商品A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「気温と商品Aの売れ行きには正の相関がある」と推測されます。



エクセルで散布図

- 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。

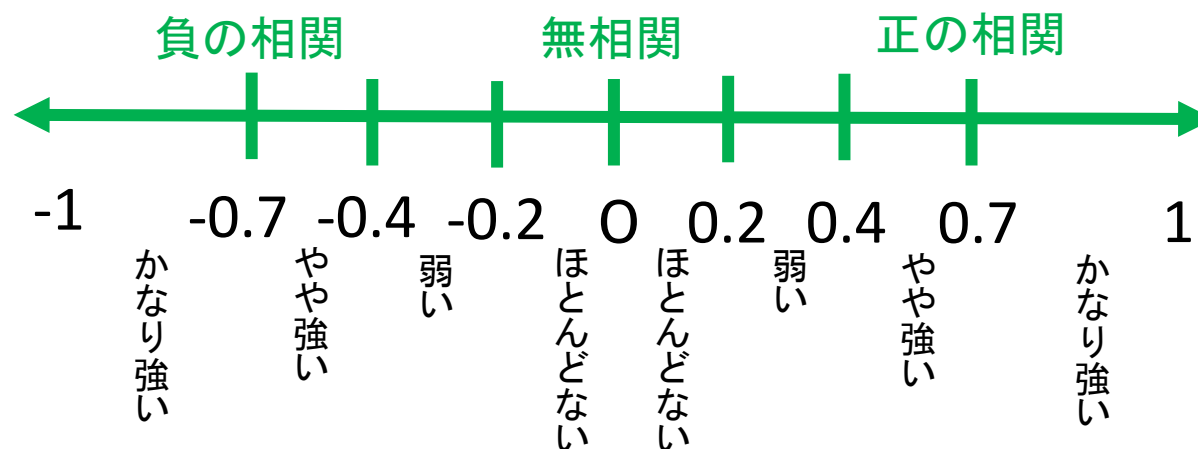


B列とC列を選択する

気温と販売数の散布図を描けた
右肩あがり！正の相関がありそうだ！

相関係数

- 散布図により正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

- 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。

Excelの「数式」タブと関数ライブラリ

関数の挿入 (fx) | オートSUM | 最近使用した関数 | 財務 | 論理 | 文字列 | 日付/時刻 | 検索/行列 | 数

関数ライブラリ

	A	B	C	D	E
1	気温と商品Aの販売数の関係				
2	日付	気温	販売数	相関係数	
3	4月1日	12.6	23		
4	4月2日	11.6	27		
5	4月3日	12.3	22		
6	4月4日	14.7	21		
7	4月5日	12.0	32		
8	4月6日	14.7	27		
9	4月7日	13.5	27		
10	4月8日	15.2	24		

関数の挿入

関数の検索(S): 相関

関数の分類(C): 候補

関数名(N):

- CORREL
- LINEST
- PEARSON
- RSQ
- COVAR
- COVARIANCE.P
- FISHERINV

CORREL(配列1,配列2)
2つの配列の相関係数を求める関数

この関数のヘルプ

OK キャンセル

エクセルで相関係数

関数の引数

CORREL

配列1 B3:B185 = {12.6;11.6;12.3;14.7;12;14.7;1...}

配列2 C3:C185 = {23;27;22;21;32;27;27;24;45;3...}

= 0.617226201

2つの配列の相関係数を返します。

配列1 には値 (数値、名前、配列、数値の範囲) を指定します。

数式の結果 = 0.617226201

[この関数のヘルプ\(H\)](#)

OK キャンセル

気温のデータの範囲




販売量のデータの範囲

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:B185,C3:C185)

気温の範囲 販売量の範囲

エクセルで相関係数

D3 :    =CORREL(B3:B185,C3:C185)

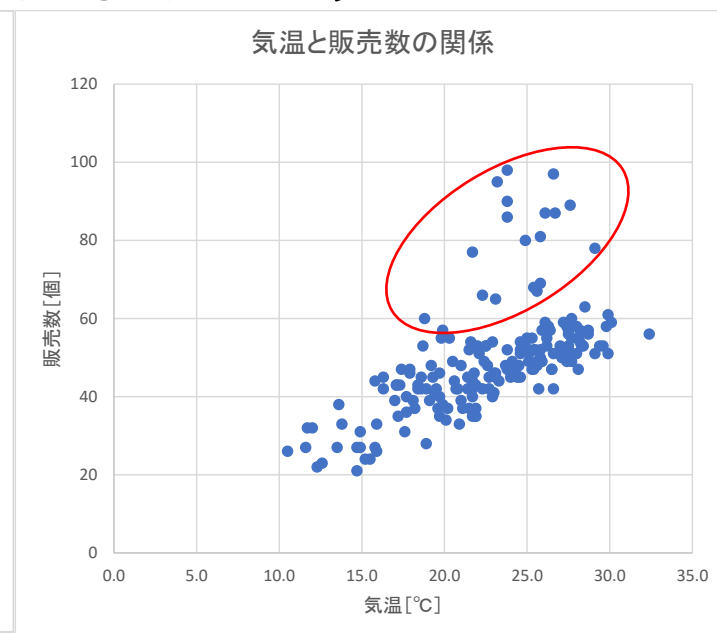
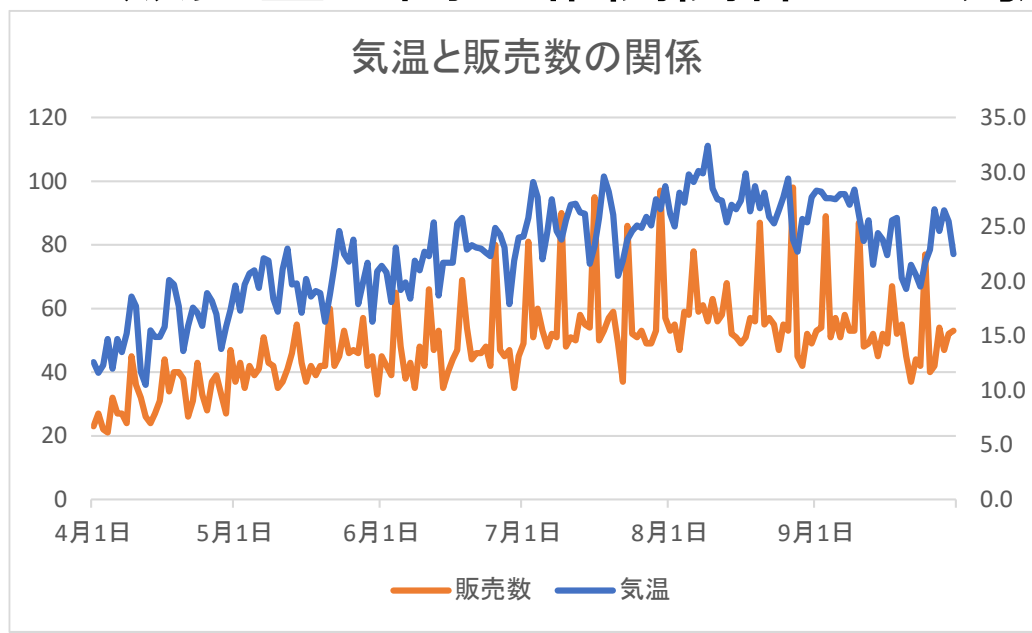
	A	B	C	D	E	F	G
1	気温と商品Aの販売数の関係						
2	日付	気温	販売数	相関係数			
3	4月1日	12.6	23	0.617226			
4	4月2日	11.6	27				
5	4月3日	12.3	22				
6	4月4日	14.7	21				
7	4月5日	12.0	32				
8	4月6日	14.7	27				
9	4月7日	13.5	27				
10	4月8日	15.2	24				
11	4月9日	18.6	45				
12	4月10日	17.7	36				
13	4月11日	11.7	32				

やや強い正
の相関がある

- 気温と商品A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！（仮説が検証された）

ここで再び仮説！

- 気温と販売量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0.61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、気温と販売量の間の相関関係はより強くなるだろうか？



データの整理

- 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)

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

E3	:	X	✓	f	=IF(WEEKDAY(\$A3,1)=1,"",B3)	
	A	B	C	D	E	F
1	気温と商品Aの販売数の関係					
2	日付	気温	販売数	相関係数	気温(除日曜)	販売数(除日曜)
3	4月1日	12.6	23	0.617226	12.6	
4	4月2日	11.6	27			
5	4月3日	12.3	22			
6	4月4日	14.7	21			
7	4月5日	12.0	32			
8	4月6日	14.7	27			
9	4月7日	13.5	27			
10	4月8日	15.2	24			
11	4月9日	18.6	45			
12	4月10日	17.7	36			
13	4月11日	11.7	32			
14	4月12日	10.5	26			
15	4月13日	15.5	24			
16	4月14日	14.9	27			

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

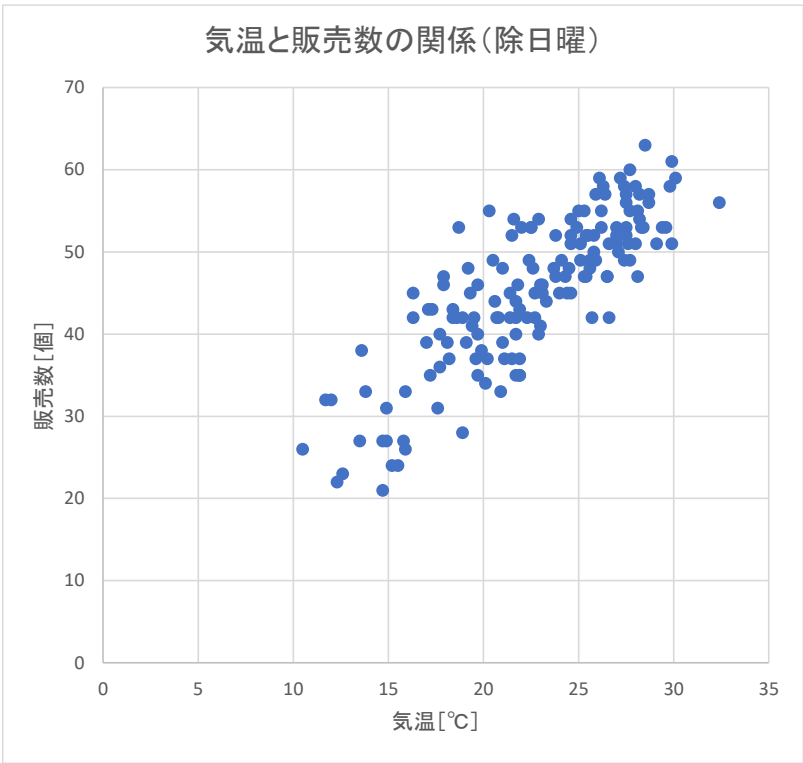
データの整理

- 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	気温と商品Aの販売数の関係					
2	日付	気温	販売数	相関係数	気温（除日曜）	販売数（除日曜）
3	4月1日	12.6	23	0.617226	12.6	23
4	4月2日	11.6	27			
5	4月3日	12.3	22		12.3	22
6	4月4日	14.7	21		14.7	21
7	4月5日	12.0	32		12	32
8	4月6日	14.7	27		14.7	27
9	4月7日	13.5	27		13.5	27
10	4月8日	15.2	24		15.2	24
11	4月9日	18.6	45			
12	4月10日	17.7	36		17.7	36
13	4月11日	11.7	32		11.7	32
14	4月12日	10.5	26		10.5	26
15	4月13日	15.5	24		15.5	24
16	4月14日	14.9	27		14.9	27
17	4月15日	14.9	31		14.9	31
18	4月16日	15.8	44			
19	4月17日	20.1	34		20.1	34
20	4月18日	19.7	40		19.7	40
21	4月19日	17.7	40		17.7	40
22	4月20日	13.6	38		13.6	38
23	4月21日	15.9	26		15.9	26
24	4月22日	17.6	31		17.6	31
25	4月23日	17.1	43			
26	4月24日	15.9	33		15.9	33

日曜日のデータは空白となった

データの再検証



E	F	G
気温 (除日曜)	販売数 (除日曜)	相関係数 (除日曜)
12.6	23	0.828188738
12.3	22	
14.7	21	
12	32	
14.7	27	
13.5	27	
15.2	24	
17.7	36	
11.7	32	
10.5	26	
15.5	24	
14.9	27	

より強い相関関係が確認できる

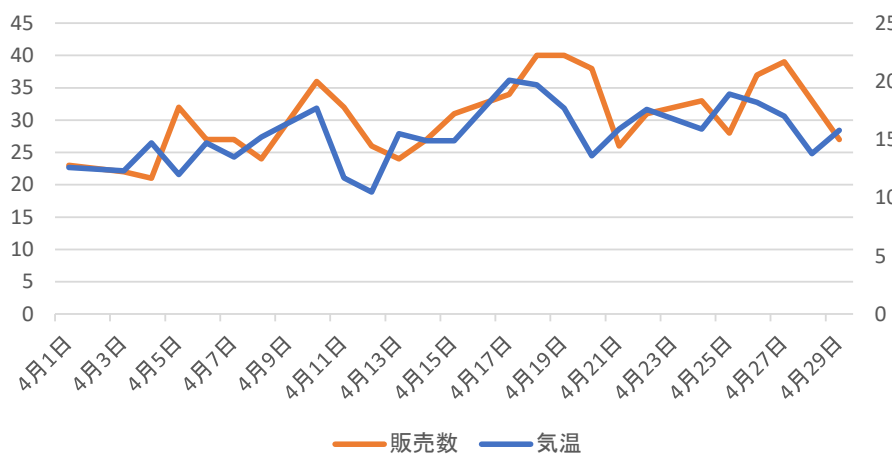
相関係数は0.828と「強い正の相関」

- 気温と日曜以外の販売量の間には強い正の相関があることが明らかとなった。

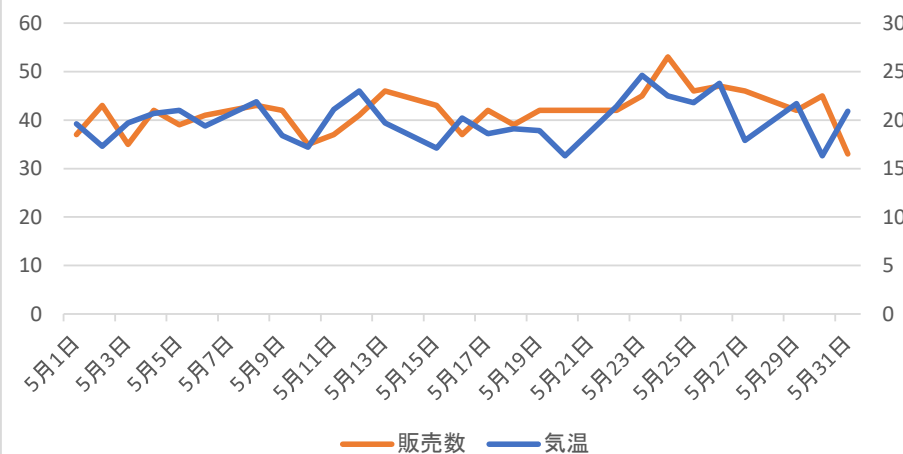
ここでもたまたま仮説！

- 実際には商品Aは1日の中でも午前中の販売量が多く、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？

4月(除日曜)



5月(除日曜)



データの整理

- まず、4月1日のセル(H3)に一日後の販売数(F4)をそのまま置く。

CORREL

X ✓ fx

=F4

	A	B	C	D	E	F	G	H
1	気温と商品Aの販売数の関係							
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	4月1日	12.6	23	0.617226	12.6	23	0.828188738	=F4
4	4月2日	11.6	27					
5	4月3日	12.3	22		12.3	22		
6	4月4日	14.7	21		14.7	21		
7	4月5日	12.0	32		12	32		
8	4月6日	14.7	27		14.7	27		
9	4月7日	13.5	27		13.5	27		
10	4月8日	15.2	24		15.2	24		
11	4月9日	18.6	45					
12	4月10日	17.7	36		17.7	36		
13	4月11日	11.7	32		11.7	32		
14	4月12日	10.5	26		10.5	26		
15	4月13日	15.5	24		15.5	24		
16	4月14日	14.9	27		14.9	27		
17	4月15日	14.9	31		14.9	31		
18	4月16日	15.8	44					
19	4月17日	20.1	34		20.1	34		
20	4月18日	19.7	40		19.7	40		

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

- 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。

H3

X

✓

fx

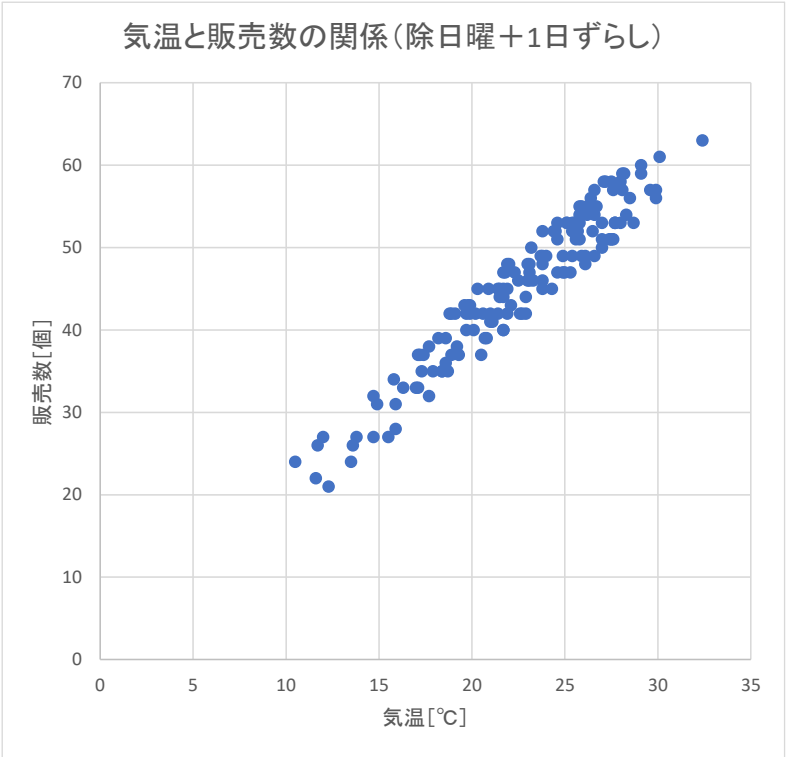
=F4

	A	B	C	D	E	F	G	H
1	気温と商品Aの販売数の関係							
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	4月1日	12.6	23	0.617226	12.6	23	0.828188738	
4	4月2日	11.6	27					22
5	4月3日	12.3	22		12.3	22		21
6	4月4日	14.7	21		14.7	21		32
7	4月5日	12.0	32		12	32		27
8	4月6日	14.7	27		14.7	27		27
9	4月7日	13.5	27		13.5	27		24
10	4月8日	15.2	24		15.2	24		
11	4月9日	18.6	45					36
12	4月10日	17.7	36		17.7	36		32
13	4月11日	11.7	32		11.7	32		26
14	4月12日	10.5	26		10.5	26		24
15	4月13日	15.5	24		15.5	24		27
16	4月14日	14.9	27		14.9	27		31
17	4月15日	14.9	31		14.9	31		
18	4月16日	15.8	44					34
19	4月17日	20.1	34		20.1	34		40
20	4月18日	19.7	40		19.7	40		40

- 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。

Aの販売数の関係		B		C		D		E		F		G		H		I	
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）										
12.6	23	0.617226	12.6	23	0.828188738												
11.6	27					22	11.6										
12.3	22		12.3	22		21	12.3										
14.7	21		14.7	21		32	14.7										
12.0	32		12	32		27	12										
14.7	27		14.7	27		27	14.7										
13.5	27		13.5	27		24	13.5										
15.2	24		15.2	24													
18.6	45					36	18.6										
17.7	36		17.7	36		32	17.7										
11.7	32		11.7	32		26	11.7										
10.5	26		10.5	26		24	10.5										
15.5	24		15.5	24		27	15.5										
14.9	27		14.9	27		31	14.9										
14.9	31		14.9	31													
15.8	44					34	15.8										
20.1	34		20.1	34		40	20.1										
19.7	40		19.7	40		40	19.7										
17.7	40		17.7	40		38	17.7										
13.6	38		13.6	38		26	13.6										

データの再々検証



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)
		0.959990992
22	11.6	
21	12.3	
32	14.7	
27	12	
27	14.7	
24	13.5	
36	18.6	
32	17.7	
26	11.7	
24	10.5	
27	15.5	
31	14.9	
34	15.8	
40	20.1	
40	19.7	

さらに強い相関関係が確認できる

相関係数は0.959と「極めて強い正の相関」

- 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正の相関があることが明らかとなった。

データ分析演習2 ～やってみよう！

- まずは, 「14時の気温と電力消費量には関係性があるのではないか？」という仮説の下で, 散布図と相関係数で検証してみましょう。
- さらに, 仮説を立ててみましょう。
 - 気温20℃以下を除いたらもっと正の相関が強くなる？
 - 土日を除いたらより相関が強くなる？
 - 祝日, 盆, 正月も除いたらもっと相関が強くなる？
- 散布図と相関係数により再度検証してみましょう。

データ分析演習2 ～もっとやってみよう！

- 余裕のある人は、以下の仮説についても検証してみよう。
 - ☐ 日最高気温と日別ピーク電力消費量にはより強い正の相関がある？
 - ☐ 日最高気温20℃以下を除いたらもっと正の相関が強くなる？
 - ☐ 土日を除いたらより相関が強くなる？
 - ☐ 祝日、盆、正月も除いたらもっと相関が強くなる？
- 散布図、相関係数により再度検証してみよう。

まとめ

- データ分析とは、仮説と検証の繰り返しにより、データの中に潜む本質にせまること。
- まずは、仮説に従って、データを適切に要約・整理します。
- データを時系列図で示すことにより、特徴や傾向を読み取ります。
- 散布図により、2つのデータの間の関係性を読み取ります。
- 相関係数により、2つのデータの間の相関関係の強さを数値で表すことができます。これにより、仮説を検証することができます。
- データ分析により、誰も気づかなかった真実を発見できると楽しいですね。そこから「新たなビジネス」が生まれます。

参考文献

- 最新エクセルのデータ分析がみるみるわかる本，秀和システム，2014.
- エクセルデータ分析超入門，羽山博著，技術評論社，2017.

記載のデータの一部は，上記参考文献のサンプルデータを利用させていただきました。ここに記して謝意を表します。