

気象データ分析チャレンジ！

回帰分析の基礎

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

- 回帰分析
- 重回帰分析

Contents

- 回帰分析
- 重回帰分析

回帰分析

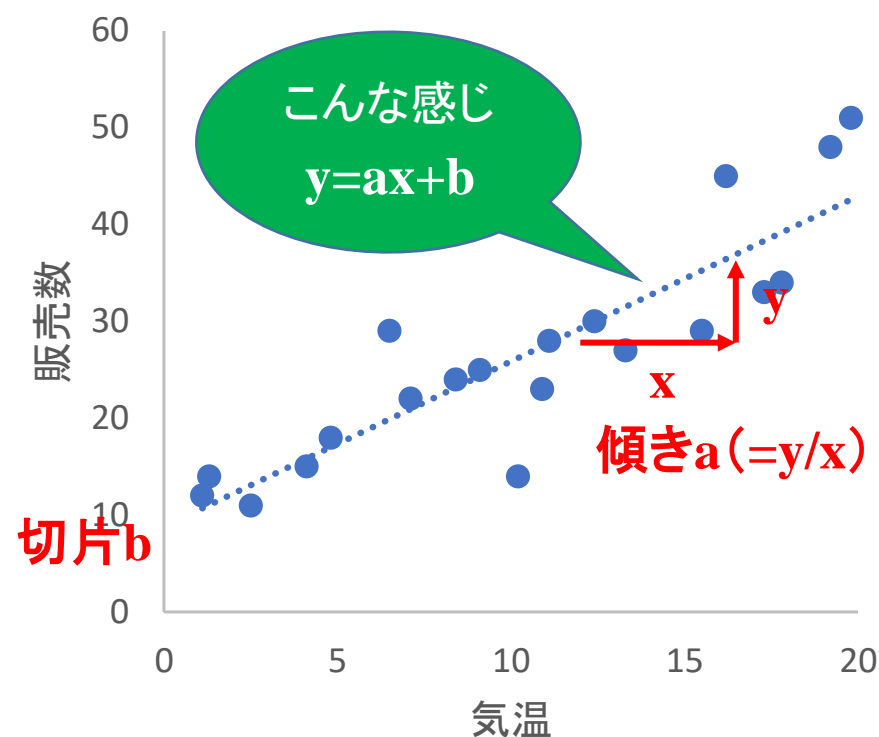
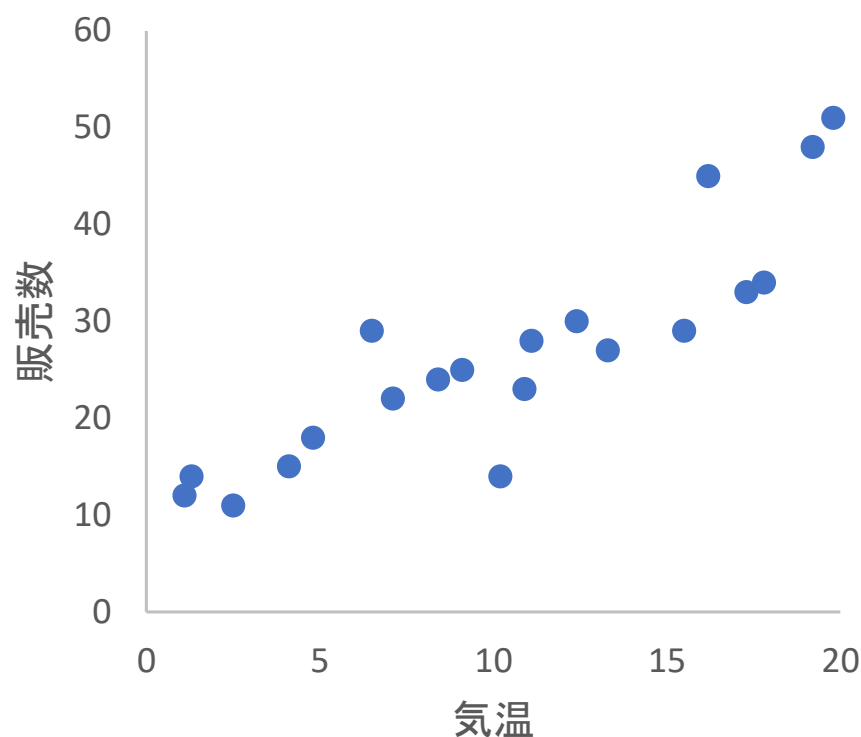
- 気温と商品Aの販売数に正の相関があることが分かったが、実務的には気温から販売数を推定できれば、仕入れや在庫の適正化を図れる。
- このような推定を行うには「**回帰分析**」という方法が使える。

$$\begin{array}{|c|} \hline \text{商品Aの} \\ \text{販売数 } y \\ \hline \end{array} = a \begin{array}{|c|} \hline \text{気温 } x \\ \hline \end{array} + b$$

気温 x (説明変数)を代入したら商品Aの販売数 y (目的変数)を推定できるような式のパラメータ a や b を知りたい！

回帰分析

- 散布図を誤差が最も小さくなるよう一本の直線で表わす。



この直線のことを**回帰直線**といい、回帰直線を表す数式を**回帰式**といいます。

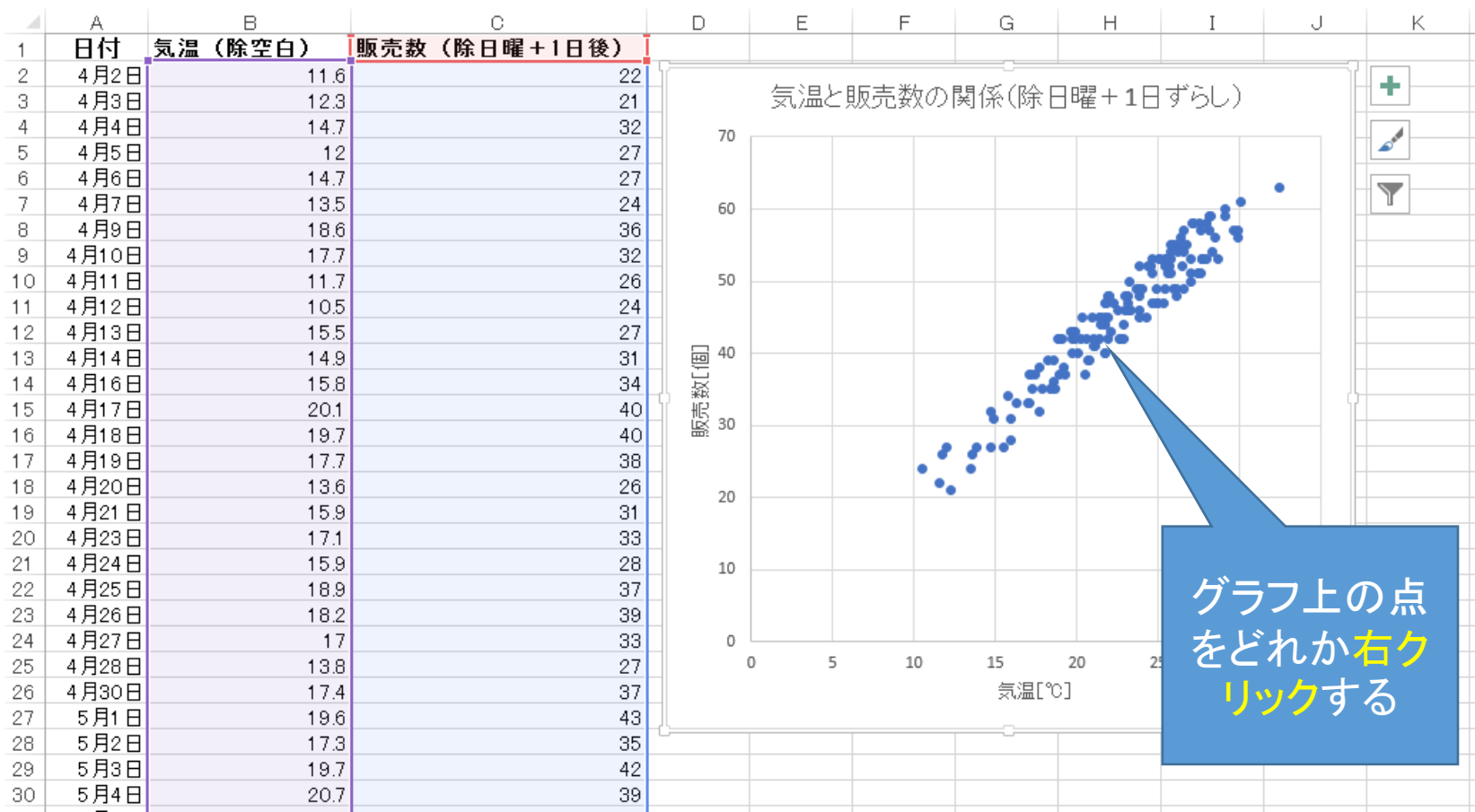
エクセルで回帰分析

- 前回の気温と商品Aの販売数(除日曜+1日ずらし)のデータ(相関係数0.959)を利用して、エクセルで回帰分析をしてみよう。

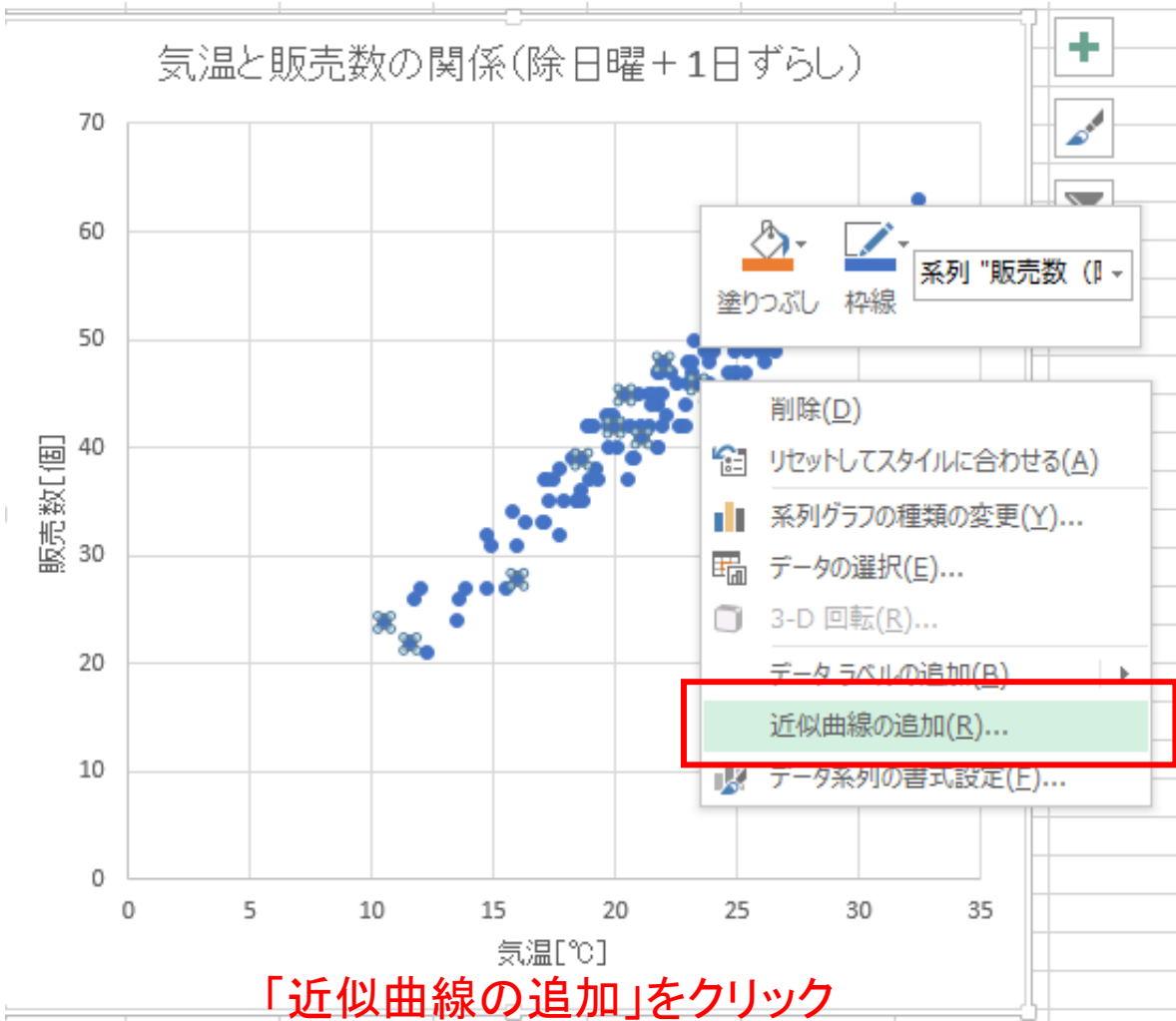
	A	B	C	D	E
1	日付	気温 (除空白)	販売数 (除日曜+1日後)		
2	4月2日	11.6	22		
3	4月3日	12.3	21		
4	4月4日	14.7	32		
5	4月5日	12	27		
6	4月6日	14.7	27		
7	4月7日	13.5	24		
8	4月9日	18.6	36		
9	4月10日	17.7	32		
10	4月11日	11.7	26		
11	4月12日	10.5	24		
12	4月13日	15.5	27		
13	4月14日	14.9	31		
14	4月16日	15.8	34		
15	4月17日	20.1	40		

散布図を使った回帰分析

- まずは前回のように散布図を作成する。



散布図を使った回帰分析



「近似曲線の追加」をクリック

近似曲線の書式設定

近似曲線のオプション ▾

近似曲線のオプション

- ☐ 指数近似(X)
- ☒ 線形近似(L)
- ☐ 対数近似(O)
- ☐ 多項式近 次数(D) 2
- ☐ 累乗近似(W)
- ☐ 移動平均(M) 区間(E) 2

近似曲線名

☒ 自動(A) 線形 (販売数 (除日曜+1日 後))

☐ ユーザー設定(C)

予測

前方補外(E) 0.0 区間

後方補外(B) 0.0 区間

☐ 切片(S) 0.0

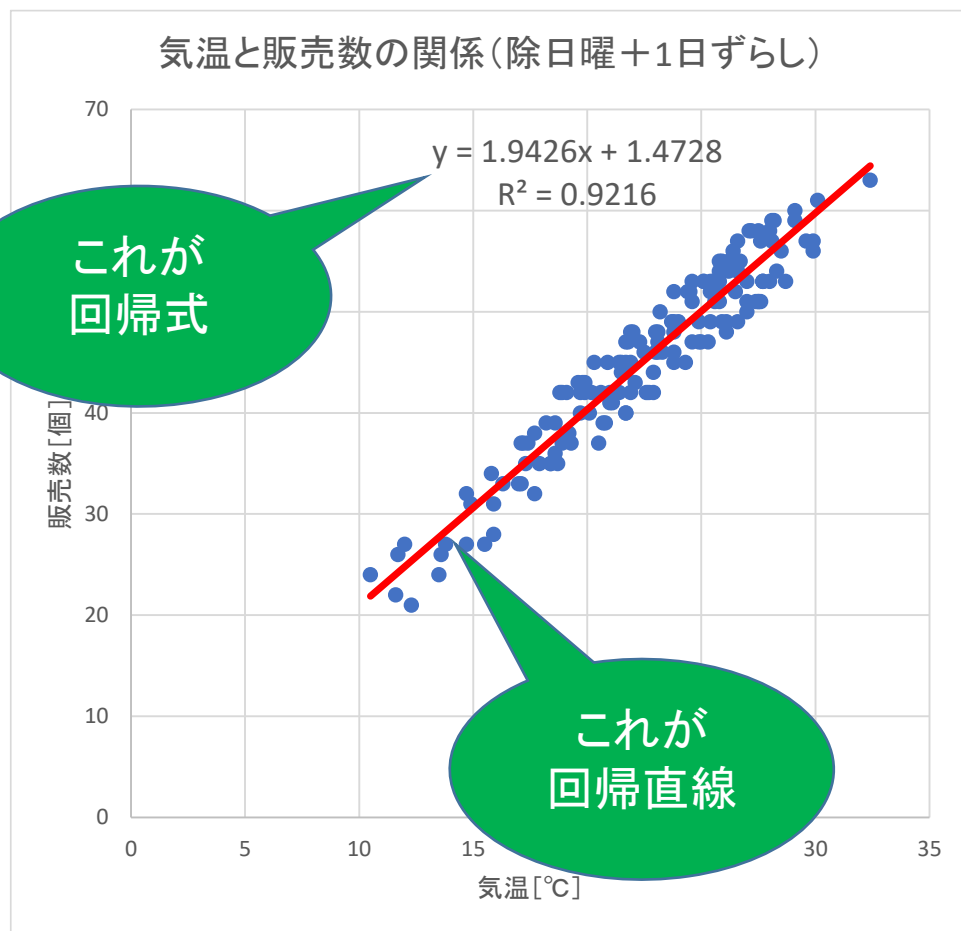
☒ グラフに数式を表示する(E)

☒ グラフに R-2 乗値を表示する(R)

「線形近似」をクリック

散布図を使った回帰分析

- 散布図中に回帰直線と回帰式を表示することができた。



回帰分析の結果, 気温と販売数の
関係を表す回帰式:

$y = 1.9426x + 1.4728$
が得られた。

傾き $a=1.94$ となった。これは, 気温
が 1°C 上昇すると1.94個多く売れる
ことを意味する。

切片 $b=1.47$ となった。これは, 気温
 0°C でも1.47個は売れることを意味
する。

関数を使った回帰分析

以下の2つの関数を利用することで回帰式 $y=ax+b$ を表現できる。

- 回帰式の傾きaを表すSLOPE関数
- 回帰式の切片bを表すINTERCEPT関数

D4				
	A	B	C	D
1	日付	気温（除空白）	販売数（除日曜+1日後）	相関係数
2	4月2日	11.6	22	0.959990992
3	4月3日	12.3	21	傾き a
4	4月4日	14.7	32	
5	4月5日	12	27	切片 b
6	4月6日	14.7	27	
7	4月7日	13.5	24	
8	4月9日	18.6	36	
9	4月10日	17.7	32	
10	4月11日	11.7	26	
11	4月12日	10.5	24	
12	4月13日	15.5	27	
13	4月14日	14.9	31	
14	4月16日	15.8	34	
15	4月17日	20.1	40	

関数を使った回帰分析

- 回帰式の傾きaの計算にはSLOPE関数を利用。

=SLOPE(C2:C157,B2:B157)

既知y(販売量)の範囲
既知x(気温)の範囲

D4				=SLOPE(C2:C157,B2:B157)	
	A	B	C	D	E
1	日付	気温 (除空白)	販売数 (除日曜+1日後)	相関係数	
2	4月2日	11.6	22	0.959990992	
3	4月3日	12.3	21	傾き a	
4	4月4日	14.7	32	1.94255508	
5	4月5日	12	27	切片 b	
6	4月6日	14.7	27		
7	4月7日	13.5	24		
8	4月9日	18.6	36		
9	4月10日	17.7	32		
10	4月11日	11.7	26		
11	4月12日	10.5	24		
12	4月13日	15.5	27		
13	4月14日	14.9	31		
14	4月16日	15.8	34		
15	4月17日	20.1	40		

傾きaは1.94
気温が1℃上昇
すると1.94個売
り上げが増える

関数を使った回帰分析

- 回帰式の切片bの計算にはINTERCEPT関数を利用。

=INTERCEPT(C2:C157,B2:B157)

既知y(販売量)の範囲
既知x(気温)の範囲

D6	:	X	✓	f_x	=INTERCEPT(C2:C157,B2:B157)
	A	B	C	D	E
1	日付	気温 (除空白)	販売数 (除日曜+1日後)	相関係数	
2	4月2日	11.6	22	0.959990992	
3	4月3日	12.3	21	傾き a	
4	4月4日	14.7	32	1.94255508	
5	4月5日	12	27	切片 b	
6	4月6日	14.7	27	1.472811832	
7	4月7日	13.5	24		
8	4月9日	18.6	36		
9	4月10日	17.7	32		
10	4月11日	11.7	26		
11	4月12日	10.5	24		
12	4月13日	15.5	27		
13	4月14日	14.9	31		
14	4月16日	15.8	34		
15	4月17日	20.1	40		

切片bは1.47
気温が0℃でも
1.47個売れる

回帰式を使った推定

- 散布図や関数から回帰式が求まれば，未知の値から結果を推定することができる。
- 気温30℃のときの販売数を推定してみよう。

= D4 * D9 + D6

傾き 気温 切片

CORREL		✕ ✓		fx	=D4*D9+D6	
	A	B	C	D	E	
1	日付	気温（除空白）	販売数（除日曜+1日後）	相関係数		
2	4月2日	11.6	22	0.959990992		
3	4月3日	12.3	21	傾き a		
4	4月4日	14.7	32	1.94255508		
5	4月5日	12	27	切片 b		
6	4月6日	14.7	27	1.472811832		
7	4月7日	13.5	24	予測		
8	4月9日	18.6	36	気温		
9	4月10日	17.7	32		販売量（推定）	
10	4月11日	11.7	26		30	=D4*D9+D6
11	4月12日	10.5	24			
12	4月13日	15.5	27			
13	4月14日	14.9	31			
14	4月16日	15.8	34			
15	4月17日	20.1	40			



回帰式を使った推定

$$=D4*D9+D6$$

C	D	E
販売数（除日曜+1日後）	相関係数	
22	0.959990992	
21	傾き a	
32	1.94255508	
27	切片 b	
27	1.472811832	
24	予測	
36	気温	販売量（推定）
32	30	59.74946424
26		
...		

- 回帰式により，気温30℃の時の翌日の販売数（除日曜）が59.7個になると推定された。

Contents

- 回帰分析
- 重回帰分析

重回帰分析

- 重回帰分析により、複数の説明変数(気温, 天気など)によって目的変数(商品Aの販売数)を予測する重回帰式を作ることができる. これによって, 決定係数 R^2 がより高くなると期待される.
- ここでは, エクセルで重回帰式を作る方法について説明します. また, 説明変数に質的データを含む場合の処理方法(数量化|類)も説明します.

商品Aの
販売数

y

=

a

気温

x_1

+

b

天気

x_2

+

c

エクセルで重回帰分析

- 気温と天気から商品Bの販売数を予測してみよう。
- 天気(晴れ, 曇り, 雨)といった数字でない情報を説明変数とする場合には, どうしたらいいのでしょうか？

	A	B	C	D	E	F
1	商品Aの販売数と気温と天気の関係					
2	販売数	気温	天気			
3	38	10	曇り			
4	27	14	曇り			
5	37	16	晴れ			
6	32	24	雨			
7	79	34	曇り			
8	42	23	雨			
9	32	12	晴れ			
10	53	21	晴れ			
11	68	25	晴れ			
12	14	10	曇り			
13	14	11	曇り			
14	22	18	雨			
15	21	10	晴れ			
16	55	17	雨			

ダミー変数への変換

- 数字でない情報を説明変数とする場合、**ダミー変数**を使って数値化します。今回の場合、晴れというダミー変数と曇りというダミー変数の2つを導入して、以下のルールでIF関数により数値を埋めます。

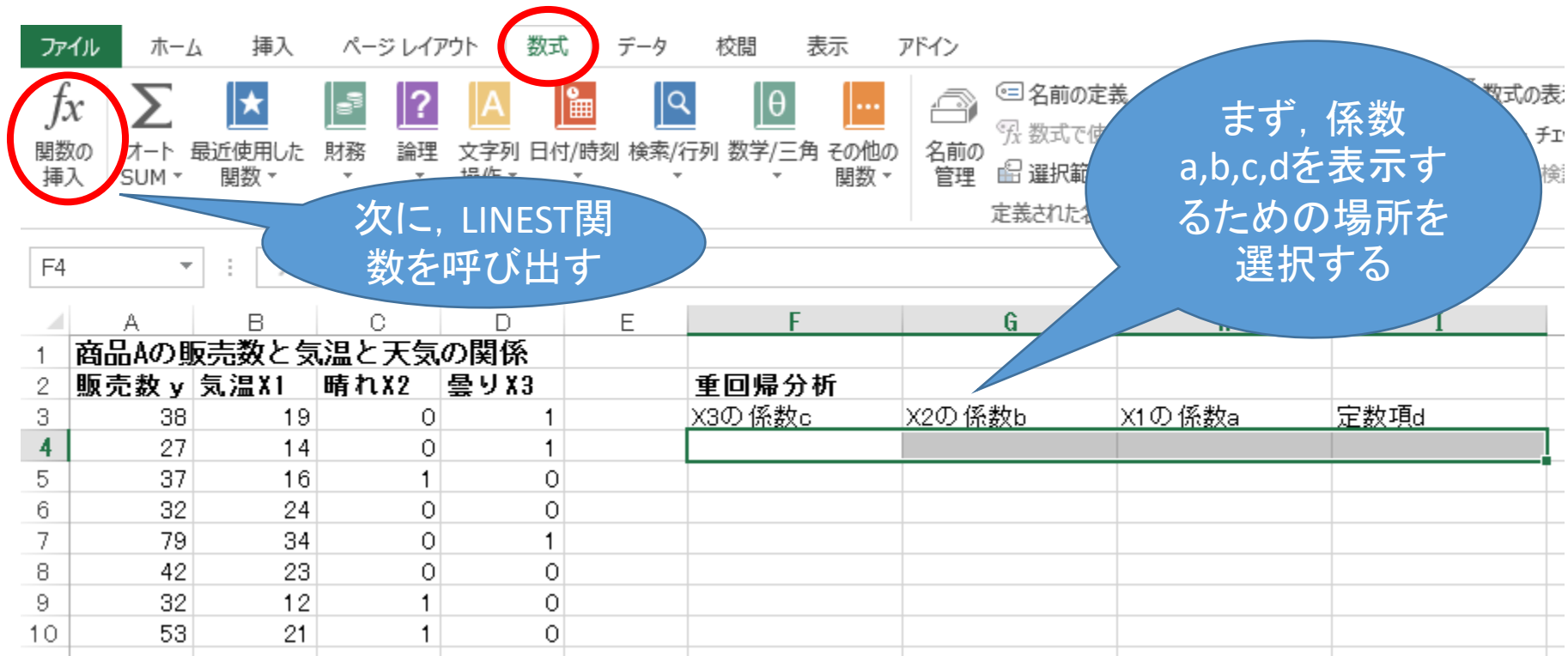
	ダミー変数 「晴れ」	ダミー変数 「曇り」
晴れ	1	0
曇り	0	1
雨	0	0

「雨」は、
晴れ=0、曇り=0
で表現する

	A	B	C	D	E	F
1	商品Aの販売数と気温と天気の関係					
2	販売数	気温	晴れ	曇り		
3	38	19	0	1		
4	27	14	0	1		
5	37	16	1	0		
6	32	24	0	0		
7	79	34	0	1		
8	42	23	0	0		
9	32	12	1	0		
10	53	21	1	0		
11	68	25	1	0		
12	14	10	0	1		
13	14	11	0	1		
14	22	18	0	0		
15	21	10	1	0		

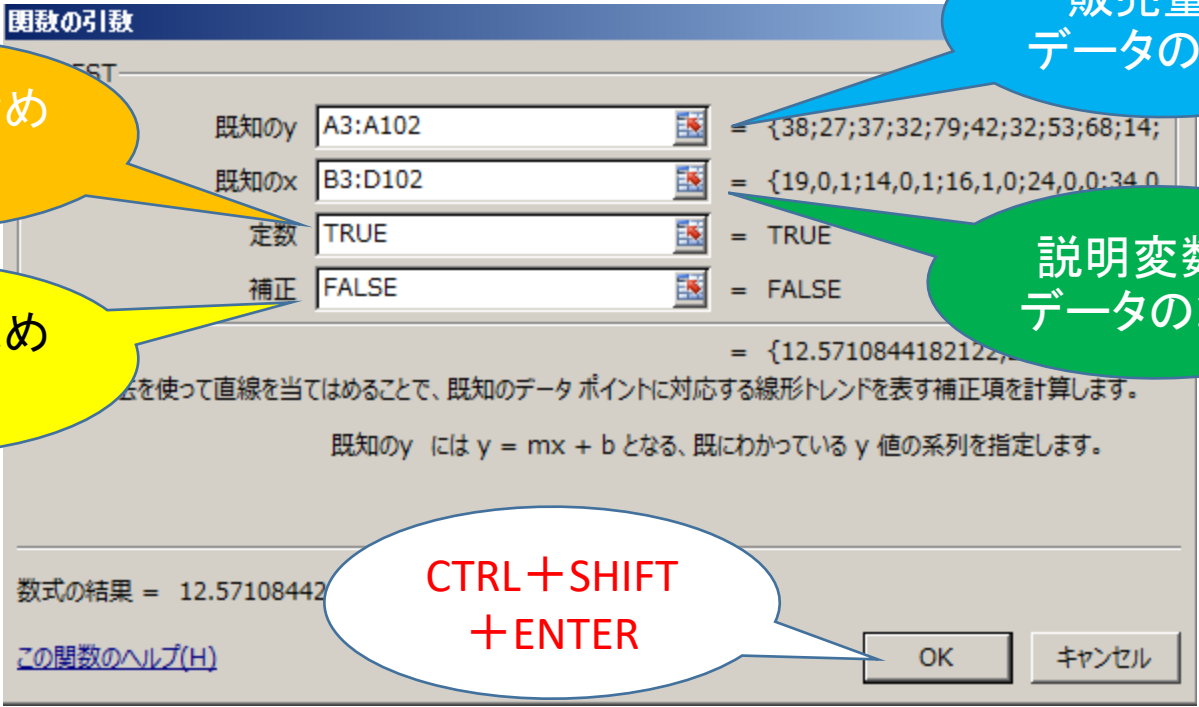
関数を使った重回帰分析

- 気温 x_1 ，晴れ x_2 ，曇り x_3 の3変数から販売量 y を推定するための重回帰式 $y=ax_1+bx_2+cx_3+d$ を導く。
- 関数「LINEST」を用いることで重回帰式を表現できる。



	A	B	C	D	E	F	G	H	I
1	商品Aの販売数と気温と天気の関係								
2	販売数 y	気温 X1	晴れ X2	曇り X3		重回帰分析			
3	38	19	0	1		X3の係数c	X2の係数b	X1の係数a	定数項d
4	27	14	0	1					
5	37	16	1	0					
6	32	24	0	0					
7	79	34	0	1					
8	42	23	0	0					
9	32	12	1	0					
10	53	21	1	0					

関数を使った重回帰分析



The image shows the '関数の引数' (Function Arguments) dialog box for the LINEST function in Excel. The dialog box has the following fields and values:

Field	Value
既知のy (Known y's)	A3:A102
既知のx (Known x's)	B3:D102
定数 (Constant)	TRUE
補正 (Regression)	FALSE

Callouts from the dialog box:

- 定数項を含める** (Include constant term): Points to the '定数' (Constant) field set to TRUE.
- 補正項を求めない** (Do not calculate correction term): Points to the '補正' (Regression) field set to FALSE.
- 販売量のデータの範囲** (Range of sales data): Points to the '既知のy' (Known y's) field.
- 説明変数のデータの範囲** (Range of explanatory variable data): Points to the '既知のx' (Known x's) field.
- CTRL+SHIFT+ENTER**: Points to the 'OK' button.

または、セルの中で関数と引数を直接打ち込む。

```
=LINEST(A3:A102,B3:D102,TRUE,FALSE)
```

販売量 の範囲	説明変数 の範囲	定数項 の有無	補正項 の有無
------------	-------------	------------	------------

関数を使った重回帰分析

F4									
1	商品Aの販売数と気温と天気の関係								
2	販売数 y	気温 X1	晴れ X2	曇り X3		重回帰分析			
3	38	19	0	1		X3の係数c	X2の係数b	X1の係数a	定数項d
4	27	14	0	1		12.57108442	23.36307907	2.932144419	-28.25619857
5	37	16	1	0					
6	32	24	0	0					
7	79	34	0	1					
8	42	23	0	0					
9	32	12	1	0					
10	53	21	1	0					
11	68	25	1	0					
12	14	10	0	1					
13	14	11	0	1					
14	22	18	0	0					
15	21	10	1	0					
16	57	27	0	0					
17	80	31	1	0					

- 気温と天気を考慮した重回帰式を作ることができた！
- 販売数 = 2.93 × (気温) + 23.36 × (晴れ) + 12.57 × (曇り) - 28.26

重回帰式を使った推定

- 重回帰式が求まれば，未知の値から結果を推定することができる。
- 気温30℃で晴れのときの販売数を推定してみよう。

$$= \underline{H4 * F7} + \underline{G4 * G7} + \underline{F4 * H7} + \underline{I4}$$

気温 (ax_1) 晴れ (bx_2) 曇り (cx_3) 切片 (d)

=H4*F7+G4*G7+F4*H7+I4

D	E	F	G	H	I	J
曇りx3		重回帰分析				
1		x3の係数c	x2の係数b	x1の係数a	定数項d	
1		12.57108442	23.36307907	2.932144419	-28.25619857	
0		予測				
0		気温	晴れ	曇り	販売数	
1		30	1	0	=H4*F7+G4*G7+F4*H7+I4	
0						
0						
0						
0						
1						
1						

重回帰式を使った推定

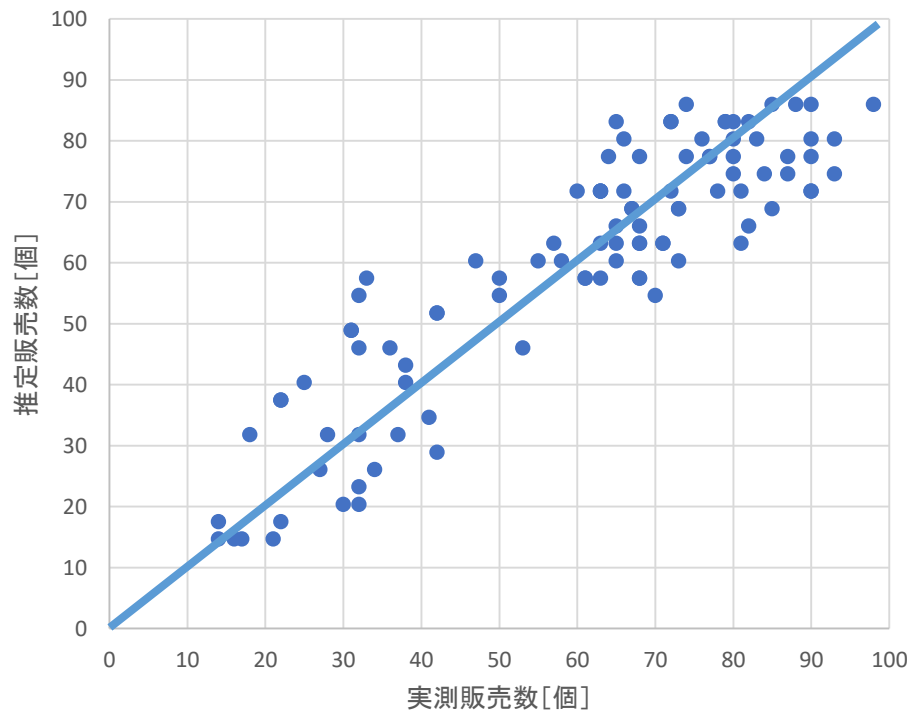
=H4*F7+G4*G7+F4*H7+I4

D	E	F	G	H	I
曇り X3		重回帰分析			
1		X3の係数c	X2の係数b	X1の係数a	定数項d
1		12.57108442	23.36307907	2.932144419	-28.25619857
0		予測			
0		気温	晴れ	曇り	販売数
1		30	1	0	83.07121307
0					
0					
0					
-					

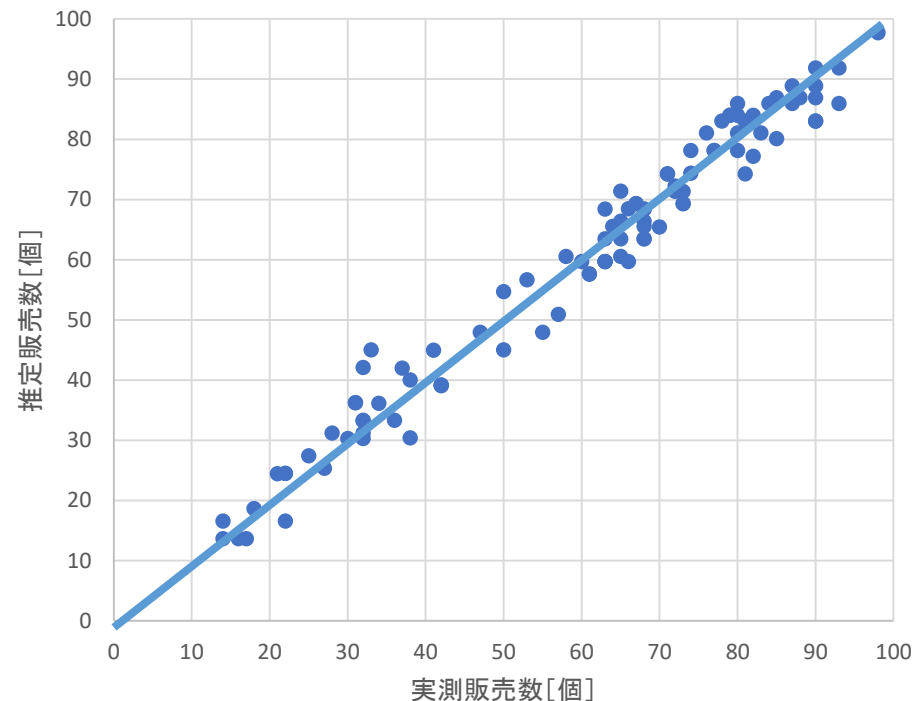
- 重回帰式により，30℃の晴れの日には83.0個の販売数になると推定された。

単回帰式vs重回帰式

推定(単回帰式)

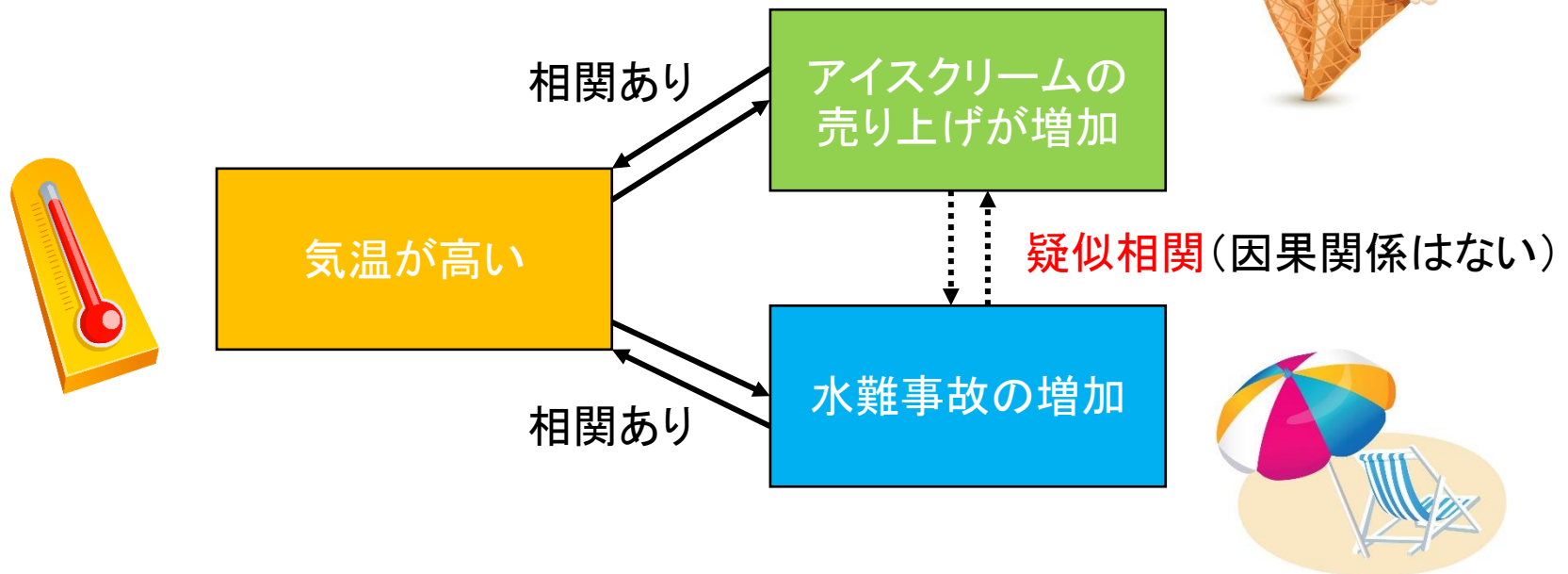


推定(重回帰式)



- 目的変数を増やすことで回帰式の精度が高くなる。
ただし、**多重共線性(説明変数の間の相関性)**に
気をつける必要がある。

擬似相関に気をつける



- 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ことがあります。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。

まとめ

- 過去のPOSデータと気象データを分析し、仮説(因果関係)を検証できたら、次は、回帰分析により販売数に関する推定式(回帰式)を作ります。
- 回帰式は説明変数を増やすことによって、精度を高めることができます。これを重回帰分析と言います。ただし、説明変数が互いに独立となるように(多重共線性を起こさないように)配慮しなくてはなりません。
- 精度の高い(R^2 の大きい)回帰式に対して、気象予報データを組み合わせることにより、販売量予測が可能となります。
- 販売量予測が可能となれば、適切な生産管理、在庫管理、販促計画を立てることができるようになります。

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。