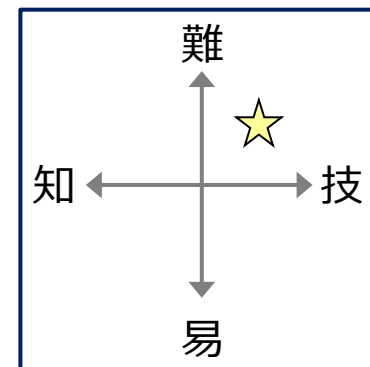
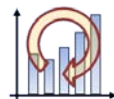


# 総務省 ICTスキル総合習得教材



[コース3] データ分析



## 3-4 : 相関と回帰分析 (最小二乗法)



[http://www.soumu.go.jp/ict\\_skill/pdf/ict\\_skill\\_3\\_4.pdf](http://www.soumu.go.jp/ict_skill/pdf/ict_skill_3_4.pdf)

	1	2	3	4	5
[コース1] データ収集					
[コース2] データ蓄積					
[コース3] データ分析				●	
[コース4] データ利活用					

# 本講座の学習内容 [3-4 : 相関と回帰分析 (最小二乗法) ]

## 【講座概要】

- Excelの散布図の作成方法、相関係数の導出方法、注意点を示します。
- 回帰分析（最小二乗法）の発想と用途を紹介します。
- Excelの分析ツールを用いた重回帰分析の実行方法を示します。
- Excelの分析ツールによる回帰分析の出力の直感的な意味を回帰分析全体と個別の説明変数に分けて説明します。

## 【講座構成】

### 実習

- [1] 散布図と相関係数
- [2] 回帰分析（最小二乗法）の発想
- [3] 分析ツールによる回帰分析
- [4] 回帰分析の全体に関する出力
- [5] 回帰分析の説明変数に関する出力

## 【学習のゴール】

- ✓ 散布図の表示、相関係数の導出方法、相関係数の利用上の注意点を把握する。
- ✓ 回帰分析（最小二乗法）の考え方を把握する。
- ✓ Excelの分析ツールを利用して、重回帰分析を行うことができる。
- ✓ Excelの回帰分析の出力の意味を理解する。

# 相関分析と回帰分析

## ◆この講座では、「相関分析」と「回帰分析」をExcelで実習します。

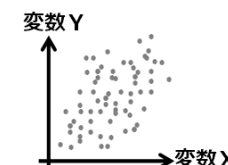
- この講座で説明する「相関分析」と「回帰分析」は、ともに連続的な数値をとる変数間の関係进行分析する手法です。

- **連続的な数値**とは、講座3-3で示した**間隔尺度**  または**比率尺度**  に対応しています。

- 名義尺度や順序尺度に相当する変数間の関係进行分析する手法もありますが、標準的に利用される分析方法や手順の範疇外となります。
- 来店客数などの人数は比率尺度ですが、「来店客が来るか？来ないか？」という来店客が0のケースがあり得るデータには、標準的な相関分析・回帰分析は適していません。一方で、来店客数の最小値が1人を上回り、来店客数の多寡进行分析する場合には、相関分析・回帰分析が適しています。

- 「相関分析」と「回帰分析」は、ともに『ある変数の大小が、他の変数の大小に結びついているか？』を分析する点では共通していますが、「分析者による因果関係の仮定」や「分析目的」に違いがあります。

----- 相関分析 ↔ 回帰分析 -----



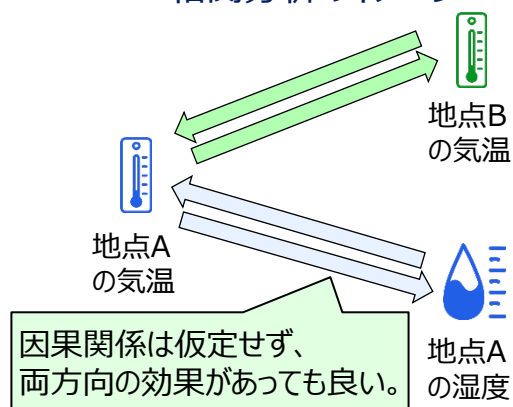
- 変数Xと変数Yの関係を分析する場合において、相関分析では因果関係を仮定しませんが、**回帰分析では「変数Xが変数Yに影響を与える」という「変数X→変数Y」の因果関係を分析者が仮定した上で分析を行います。**
- **相関分析では2種類の変数を採りあげ、変数間の（直線的な）関係の強さを測定します。**
- 回帰分析の「変数X→変数Y」の因果関係において、**矢印の元になっている変数Xを説明変数**といい、**矢印の先になっている変数Yを被説明変数**といいます。
  - 回帰分析においては、説明変数Xには、XA、XB、XC…と複数の種類があっても同時に分析に入れることができます。
  - 回帰分析の説明変数は独立変数とも呼ばれ、被説明変数は従属変数や目的変数とも呼ばれます。
- 回帰分析では、**説明変数が被説明変数に与える効果の測定を行ったり、回帰分析の結果に基づいて設定した状況や将来に関する予測を行ったりします。**

# 相関分析と回帰分析が取り扱える変数の数

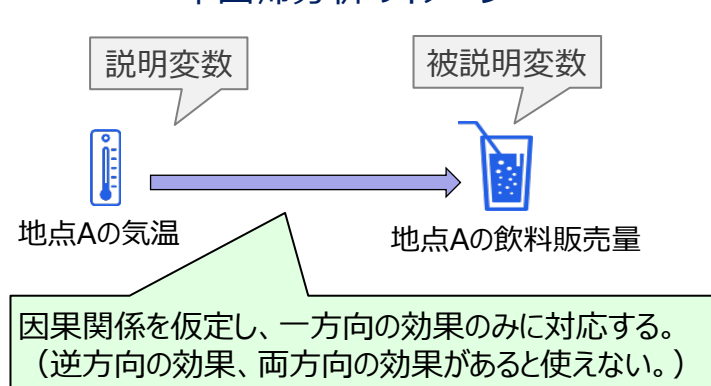
◆相関分析と回帰分析では、同時に扱うことができる変数の種類の数異なります。

- 相関分析は、因果関係を仮定せず、常に一対一の組み合わせで変数同士の（直線的な）関係を分析します。
  - ・ 相関行列として複数の相関係数をまとめて表示するケースもありますが、その要素となる各相関係数はそれぞれ一対一の変数の関係を示しています。
- 回帰分析では、分析者が説明変数（複数種も可）から被説明変数への因果関係を仮定して分析を行います。
- 回帰分析では説明変数が1種の場合を単回帰分析といい、説明変数が2種以上の場合を重回帰分析といいます。

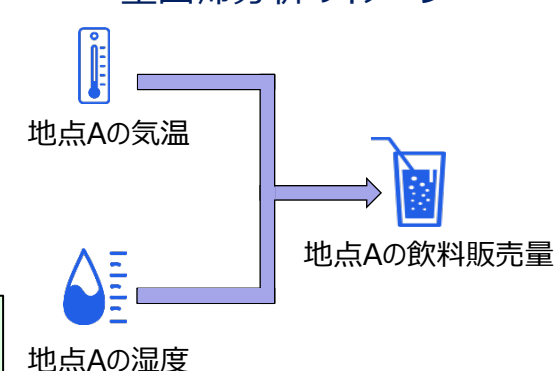
相関分析のイメージ



単回帰分析のイメージ



重回帰分析のイメージ



相関分析・回帰分析における仮定・分析目的・扱う変数の種類数の違い

分析の種類	分析者による因果関係の仮定	分析目的	扱う変数の種類数
相関分析	なし	<ul style="list-style-type: none"> <li>変数間の（直線的な）関係の強さの測定</li> </ul>	2種類（一対一）
回帰分析	あり	<ul style="list-style-type: none"> <li>説明変数が被説明変数へ与える効果の測定</li> <li>回帰分析の結果に基づく予測</li> </ul>	2種類以上 (説明変数は複数種可)

- この講座では「【実習用データ】ICT3-4\_相関と回帰分析（最小二乗法）.xlsx」を用いて実習を行います。  
利用するExcelのシート番号は、各スライド右上の〔〕内に示します。 [http://www.soumu.go.jp/ict\\_skill/dc/ict\\_3\\_4data.zip](http://www.soumu.go.jp/ict_skill/dc/ict_3_4data.zip)

# Excelの「分析ツール」の導入

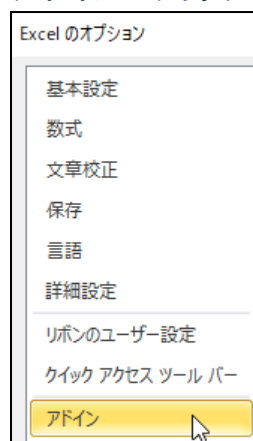
◆本講座ではExcelの分析ツールを利用するため、事前にアドインから導入しておきます。

- 相関分析・回帰分析の実習ではExcelの**分析ツール**を利用するため、未導入の場合は**アドインから導入**します。

🖥️ Excel画面の左上にある「ファイル」のボタンをクリックした後、「オプション」⇒「アドイン」をクリックして、「Excelアドイン」の設定ボタンをクリック後に「分析ツール」にチェックを入れて「OK」をクリックします。

- Excelのバージョンによって、分析ツールのアドインを呼び出す項目が異なります。ここではWindows用のExcel 2010で例示をしています。
- Macintosh用の「Excel for Mac 2011」以前では分析ツールのアドインはありません。「Excel for Mac 2016」から、Macintosh用のExcelにおいても分析ツールのアドインが利用できるようになりました。

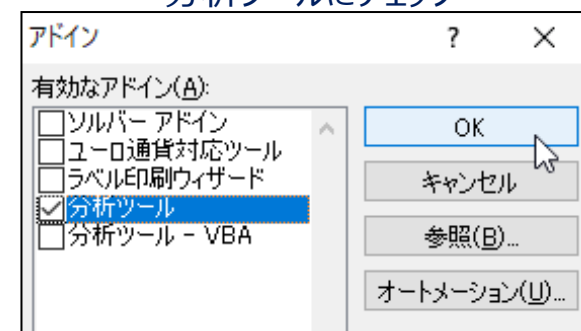
アドインのクリック



Excelアドインの設定ボタンをクリック



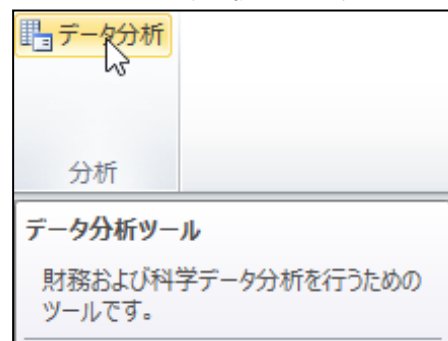
分析ツールにチェック



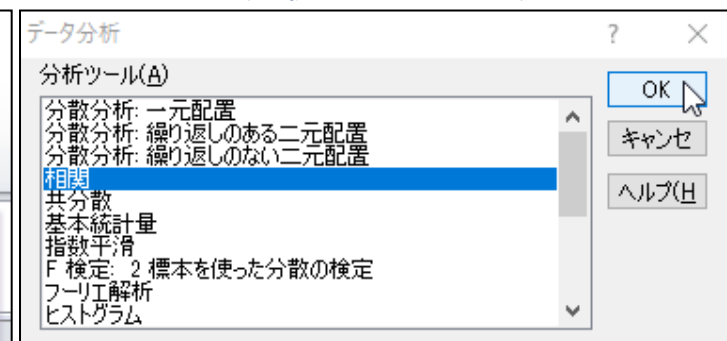
🖥️ 「分析ツール」の導入後、Excel画面の上側の「データ」タブに現れた「データ分析」をクリックすることで「データ分析」のダイアログボックスを呼び出すことができます。

- 講座内の2箇所ではExcelの「分析ツール」を利用して分析を行います。

データ分析をクリック



データ分析のダイアログボックス



# 散布図の作成

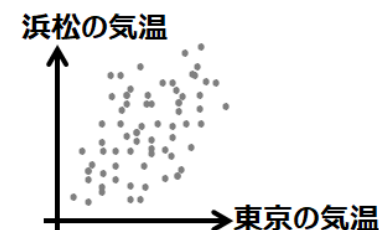
## ◆実習用データから2種類の変数を指定して散布図を描きます。

- 実習用データのシート〔1〕には、東京と浜松における日平均気温・湿度および飲料販売量が記入されています。

- 気温、湿度のデータは気象庁のウェブサイトからダウンロードした2016年に関する現実のデータ、飲料販売量は架空のデータです。

【出所】過去の気象データ・ダウンロード【気象庁】 <http://www.data.jma.go.jp/gmd/risk/obsdl/index.php>

- ここでは散布図作成の例示するために〔東京：平均気温(℃)〕と〔浜松：平均気温(℃)〕の2変数の関係を描くことを考えます。



- 散布図の対象データとする2列をまとめて選択した状態で、Excelの上部の「挿入」タブにある「散布図」ボタンをクリックし、その中にある「マーカーのみ」のボタンをクリックします。

「挿入」タブにある「散布図」から「散布図（マーカーのみ）のボタン」をクリック

東京: 平均気温(℃)

A	B	C	D
日付	曜日(休日)	東京: 平均気温(℃)	浜松: 平均気温(℃)
2016/1/1	金(休)	7.5	7.7
2016/1/2	土(休)	7.3	8.5

散布図 (マーカーのみ)  
値の組を比較します。



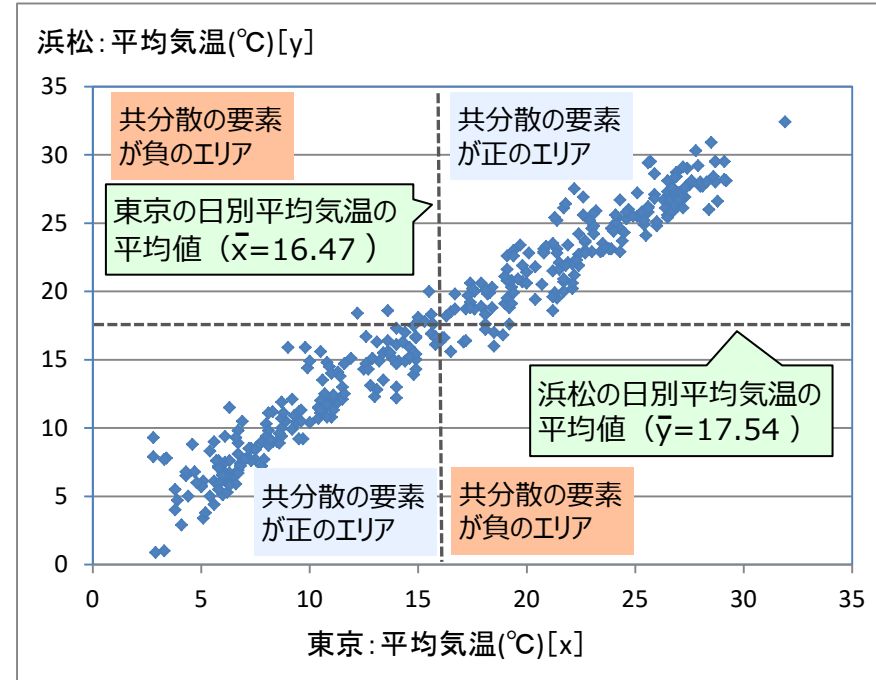
# 散布図と相関係数

## ◆相関係数は、2種類の連続的な変数の関係を一つの値で要約します。

- 作成した散布図は右図のように、東京の気温が高い場合に浜松の気温も高い傾向があり、直線的な関係が確認できます。
- 一方の変数が大きいと、もう一方の変数も大きい傾向があり、右肩上がりの直線的な関係が確認できる場合は**正の相関関係**にあるといいます。
- 直線的な関係の傾きが右肩下がり場合は負の相関関係にあるといい、直線的な関係が真横であったり、直線的な関係が確認しにくい場合は無相関といいます。
- 相関関係を「-1～1」の範囲で一つの値で要約する指標が、**相関係数**であり、下記の式で導出できます。
- 右図では東京の気温をx、浜松を気温をyと取り扱っています。

$$x \text{ と } y \text{ の相関係数} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

東京と浜松の気温の散布図（2016年：日別平均）



- 相関係数の分子は、変数xと変数yの**共分散**（きょうぶんさん）と呼ばれます。共分散は各標本の「『変数xの偏差』と『変数yの偏差』の積」を要素として、標本での総和をとってから標本数で割ること（⇒平均値）で算出できます。
- 右上のグラフにおいて、それぞれの平均値を軸として、右上と左下のエリアの要素は正の値となり、左上、右下のエリアの要素は負の値となります。
- 相関係数の分母は変数x、変数yの各標準偏差となっており、相関係数を「-1～1」の間に調整する役割を持ちます。
- 標準偏差は、講座3-3で示したように変数のバラツキの指標として、二乗した値の総和によって導出しているため、必ず正の値になります。
- 相関係数は、**正と負の値をとりうる共分散が分子、正の値のみをとる「-1～1」への調整項が分母**になっています。

# Excel関数CORRELによる相関係数の導出

## ◆Excel関数CORRELで2種類の変数を指定して相関係数を導出することができます。

- 相関係数の分子の**共分散**は、**各変数の偏差の積を要素とする平均値**となっていることから、散布図における各変数の平均値の軸から「右上」と「左下」が多ければ正の値、「右下」と「左上」が多ければ負の値になります。
- Excel関数「**COVARIANCE.P**」を利用すると、（母集団としての）共分散が導出できます。
  - 講座3-3において分散や標準偏差で示したように、標本数(n)ではなく(n-1)で割ることで標本共分散を導出するExcel関数「COVARIANCE.S」もあります。なお、分子を標本共分散とする場合は、対応する分母も標本標準偏差となり、分子分母をともに(n-1)で割るため、相関係数は一致します。
- 変数x、変数yのそれぞれの（母集団としての）標準偏差は、講座3-3で示したように「**STDEV.P**」で導出できます。
- 変数xと変数yの共分散をそれぞれの変数の標準偏差で割ることによっても、Excel関数「**CORREL**」を利用することによっても、変数xと変数yの相関係数を導出することができます。
- [東京：平均気温(℃)] と [浜松：平均気温(℃)] の相関係数として0.977が導出できます。



実習用データ（浜松の気温・東京の気温）に関する  
共分散、相関係数の出力 [シート〔2〕のL列]

指標	Excel関数の入力	出力
xの平均値	=AVERAGE(C2:C367)	16.47
yの平均値	=AVERAGE(D2:D367)	17.54
xとyの共分散	=COVARIANCE.P(C2:C367,D2:D367)	55.997
xの標準偏差	=STDEV.P(C2:C367)	7.668
yの標準偏差	=STDEV.P(D2:D367)	7.474
xとyの相関係数	=CORREL(C2:C367,D2:D367)	0.977

共分散を分子、  
各標準偏差の積を分母  
とする割り算

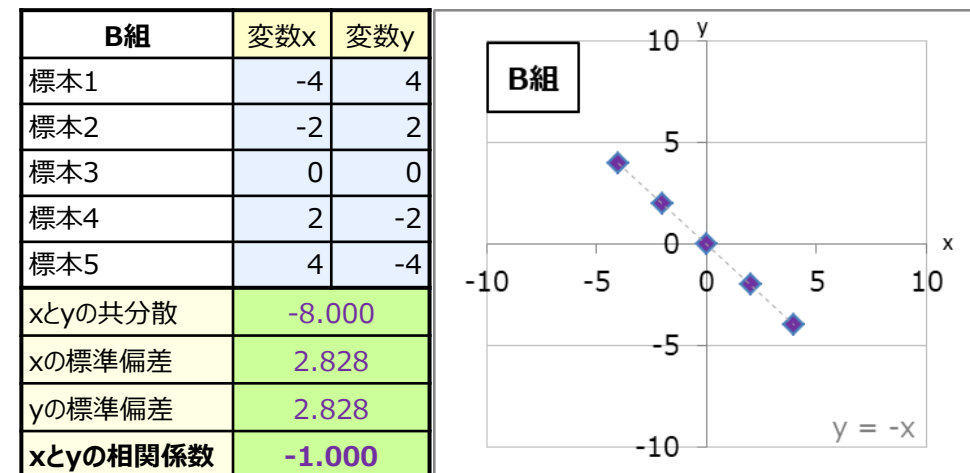
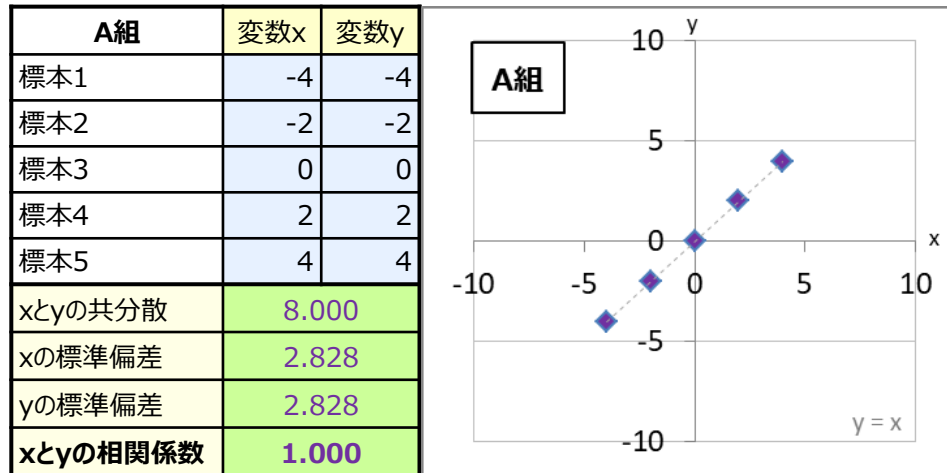
□ 続く2枚のスライドでは、5つの標本での簡潔な数値例で、相関係数の特徴を紹介します。



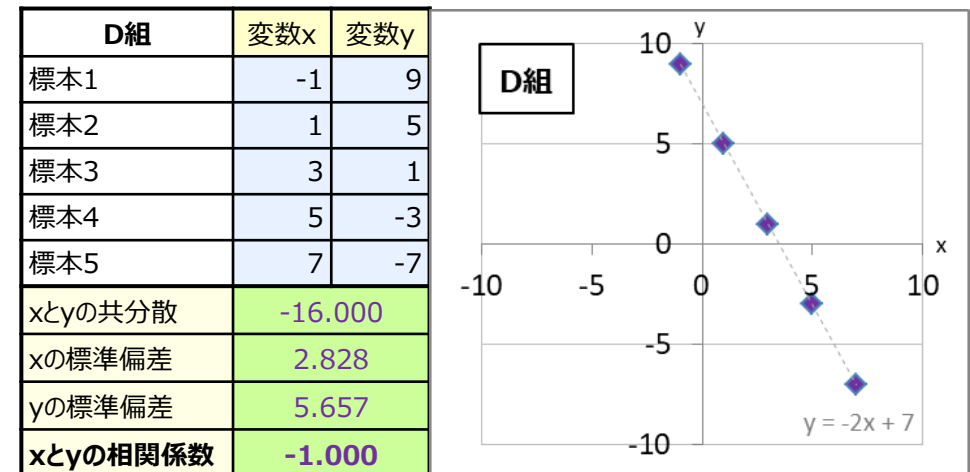
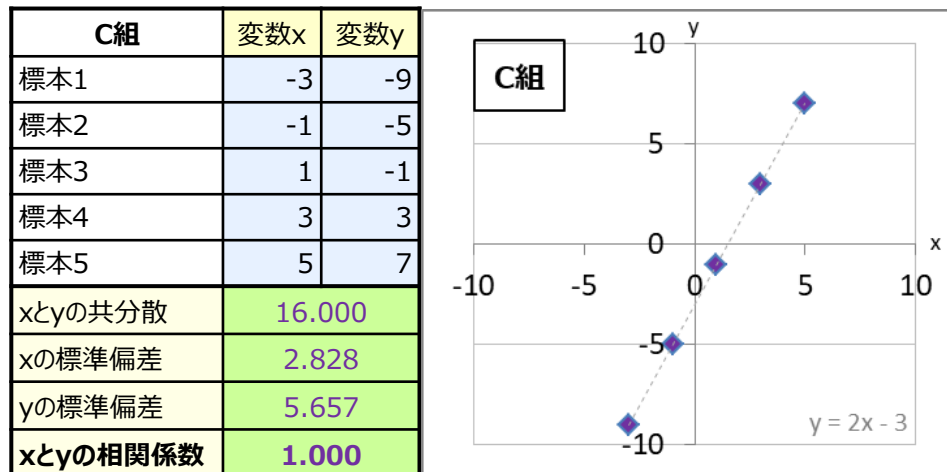
# 【参考】相関係数と散布図の例示（1）

## ◆変数間の関係を示す直線上に全標本が乗っている場合、相関係数は絶対値で1になります。

- 相関係数は、標本全てが右肩上がりの直線上にあれば+1、標本全てが右肩下がりの直線上にあれば-1となります。
  - 散布図の右下には、後述する「散布図における単回帰分析」を行った場合に表示される回帰線の数式を示しています。



- 標本を平行移動させても、（全標本が直線上にある場合は）直線の傾きが変わっても、相関係数は変わりません。

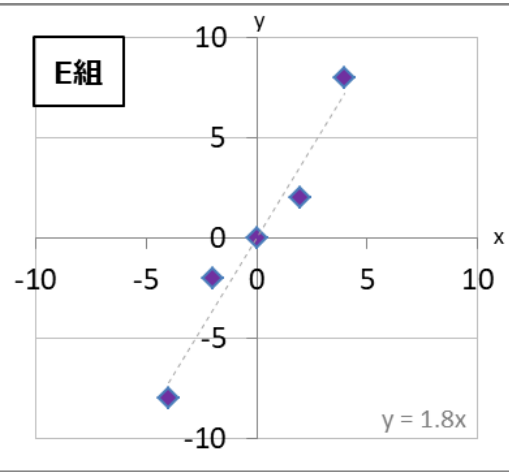


# 【参考】相関係数と散布図の例示（2）

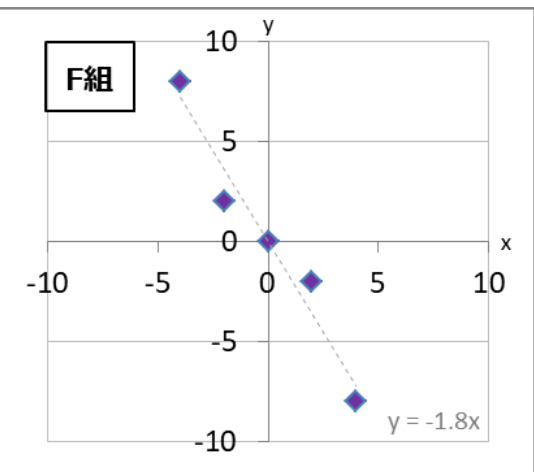
◆変数間の関係を示す直線から外れる標本があれば、相関係数は絶対値で1を下回ります。

- 変数間の変化が一方向でも、直線的な関係から外れる標本があれば、相関係数は絶対値で1を下回ります。

E組	変数x	変数y
標本1	-4	-8
標本2	-2	-2
標本3	0	0
標本4	2	2
標本5	4	8
xとyの共分散	14.400	
xの標準偏差	2.828	
yの標準偏差	5.215	
xとyの相関係数	0.976	

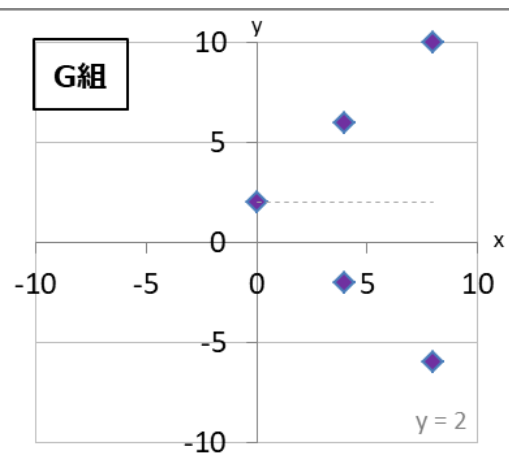


F組	変数x	変数y
標本1	-4	8
標本2	-2	2
標本3	0	0
標本4	2	-2
標本5	4	-8
xとyの共分散	-14.400	
xの標準偏差	2.828	
yの標準偏差	5.215	
xとyの相関係数	-0.976	

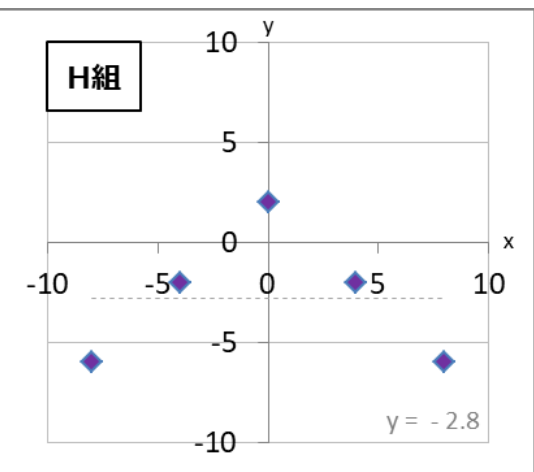


- 相関係数は屈曲した関係を反映せず、2種類の変数間に関係があっても、相関係数が0になるケースもあります。

G組	変数x	変数y
標本1	8	10
標本2	4	6
標本3	0	2
標本4	4	-2
標本5	8	-6
xとyの共分散	0.000	
xの標準偏差	2.993	
yの標準偏差	5.657	
xとyの相関係数	0.000	



H組	変数x	変数y
標本1	-8	-6
標本2	-4	-2
標本3	0	2
標本4	4	-2
標本5	8	-6
xとyの共分散	0.000	
xの標準偏差	5.657	
yの標準偏差	2.993	
xとyの相関係数	0.000	



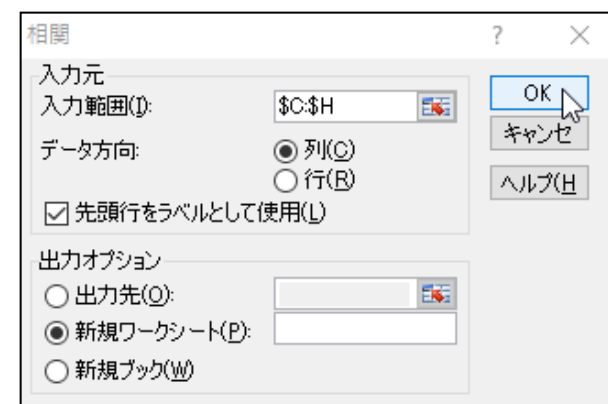
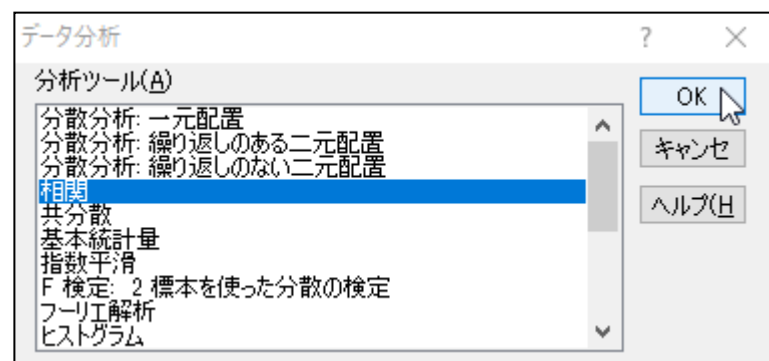
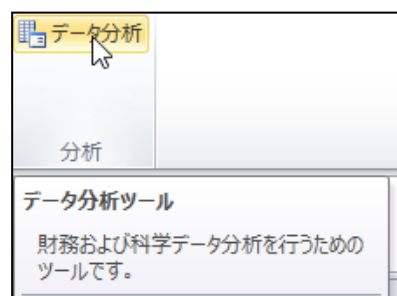
□ シート[参1]では、散布図における各標本の位置を変更して、相関係数がどのように変化するかを確認できます。

# 分析ツールによる相関係数の一括導出（相関行列）

## ◆Excelの分析ツールから、様々な変数の組み合わせの相関係数を一括して導出できます。

- Excelの「分析ツール」の「相関」から、相関係数の組み合わせを一括表示する相関行列を作成できます。

Excel画面上側の「データ」タブの「データ分析」をクリックし、ダイアログボックスから「相関」を選択し、「OK」をクリックします。  
 入力範囲としてシート[1]のC列の「東京：平均気温(℃)」～H列「浜松店：飲料販売量(本)」を選択します。  
 「先頭行をラベルとして使用」にチェックを入れ「新規ワークシート」を選択し、「OK」をクリックします。



- 出力された相関行列では、縦の列と横の行の交差する欄に各2変数の相関係数が記入されています。
- Excel関数で導出した相関係数を確認できることに加え、シート[1]にある変数は全ての組み合わせで正の相関関係にあることが分かります。

### 分析ツールの「相関」による相関行列の出力

	東京： 平均気温(℃)	浜松： 平均気温(℃)	東京： 平均湿度(%)	浜松： 平均湿度(%)	東京店：飲料 販売量(本)	浜松店：飲料 販売量(本)
東京：平均気温(℃)	1					
浜松：平均気温(℃)	0.977	1				
東京：平均湿度(%)	0.493	0.589	1			
浜松：平均湿度(%)	0.525	0.561	0.809	1		
東京店：飲料販売量(本)	0.548	0.543	0.365	0.367	1	
浜松店：飲料販売量(本)	0.671	0.702	0.554	0.596	0.644	1

- 相関行列は、様々なパターンの2変数の組み合わせの直線的な関係を、一覧で確認したい場合に便利です。

# 散布図における単回帰分析

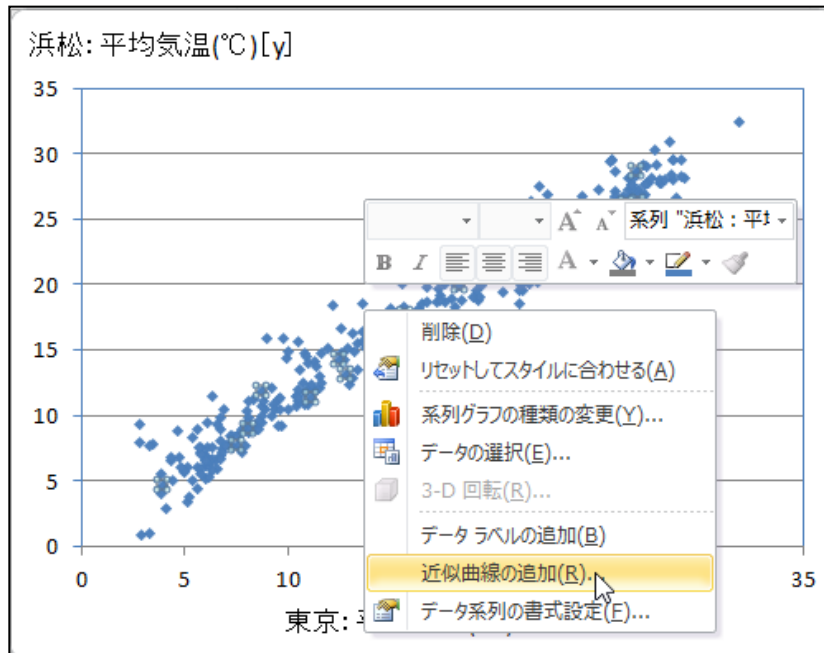
## ◆Excelの散布図においても、単回帰分析の結果を確認することができます。

- Excelの散布図においても、説明変数が1種類の単回帰分析の結果を確認することができます。

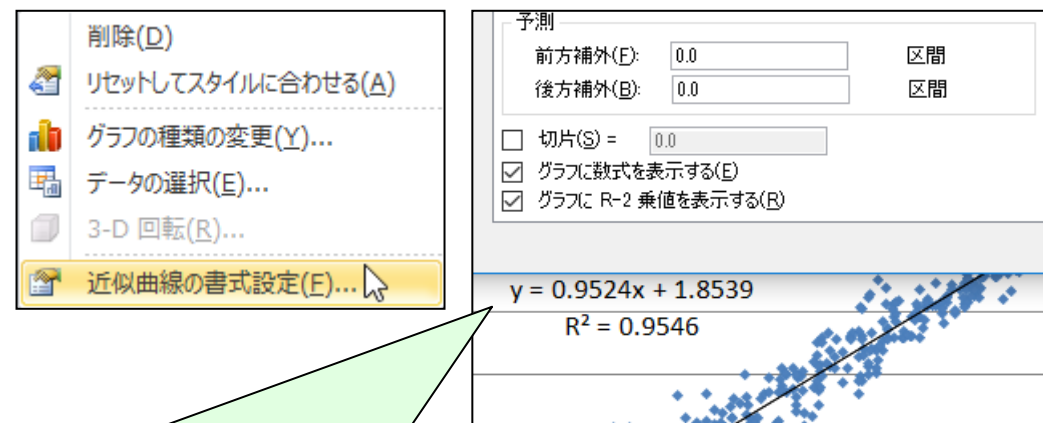
🖱 散布図を右クリックして表示されるメニューから「近似曲線の追加」をクリックすることで、二変数の関係を示す回帰線を引きます。  
表示されたダイアログボックスから「グラフに数式を表示する」「グラフにR-2乗値を表示する」にチェックを入れてください。

- ダイアログボックスが表示されずに近似曲線が引かれた場合は、近似曲線を右クリックして「近似曲線の書式設定」から選択することができます。
- 回帰分析では「変数xが1大きくなった場合に、変数yにどの程度の影響を与えるか？」を測定することができます。
  - 同時に表示される $R^2$ は、決定係数と呼ばれる「0～1」の間をとる回帰分析の当てはまりの指標であり、後述します。
  - ここでは考察の対象外としますが、暗黙裡に「変数x [東京の気温(℃)] → 変数y [浜松の気温(℃)]」の一方向の効果を仮定しています。

### 右クリックメニューから「近似曲線の追加」



### 「近似曲線の書式設定」から「グラフに数式を表示する」にチェック



散布図上に「東京の気温と浜松の気温の関係を表す数式」が表示されます。  
この画面の結果では、

- 東京の気温が1℃増加したら、浜松の気温が0.9524℃増加する関係にあること
- 東京の気温0℃に対応する浜松の気温は1.8539℃であることを示しています。

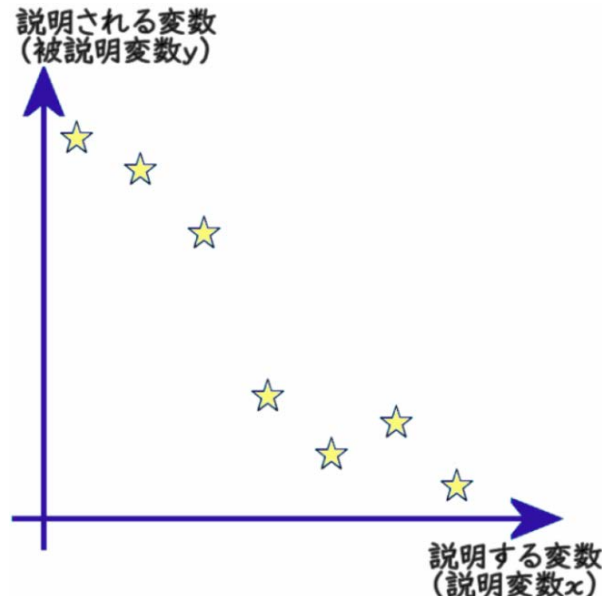
□ 続いて、回帰分析（最小二乗法）が「どのような発想で直線を引いているのか？」を紹介します。

# 回帰分析（最小二乗法）における回帰線導出のルール

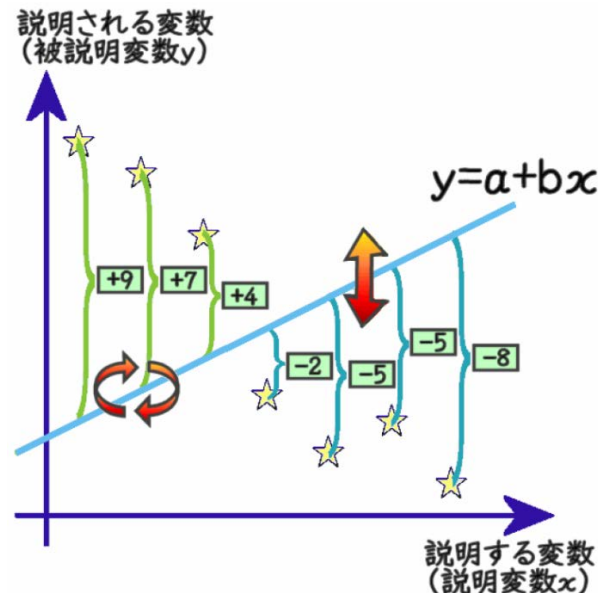
◆回帰分析は、差分を2乗の総和を最小化するルールで当てはまりの良い直線を引きます。

- 左下図のような説明される変数（被説明変数） $y$ と説明する変数（説明変数） $x$ の関係による点の星★があり、「2種類の変数の関係を示す当てはまりの良い直線（ $y=a+bx$ ）を引く」ためのルールを考えます。
  - $a$ の値を変更することで直線の高さを変えたり、 $b$ の値を変更することで直線の傾きを変えて、当てはまりの良い直線を引くことを考えます。
- 「星と直線の差分（距離）の総和を最小化するルール」が良さそうに思いますが、正と負の差分が相殺してしまいます。
  - 点のはるか上に線を引くことで、絶対値が大きい負の差分を得ることができ、差分の総和を負の値にすることも可能です。
- 「星と直線の差分を二乗して、全て正の値にしてから総和を最小化するルール」なら、負の差分の問題がありません。
- 回帰分析は、「星と直線の差分を2乗してから総和を最小化するルール」で直線（⇒回帰線）を引きます。

★ に表される「 $x$ と $y$ の関係を示すための当てはまりの良い直線」を引くためのルールを考えたい。



「★と直線の差分の総和を最小化」するルールでは、正と負の差分で相殺し、当てはまりの良い直線にならない。



★と直線の差分を2乗してから、「『全て正の値に変換した差分』の総和を最小化するルール」なら、当てはまりの良い直線が引ける。

2乗した差分の総和を最小化するため、**回帰分析は最小二乗法**とも呼ばれます。

- 最小二乗法の方が分析における利用頻度は高いですが、「星と直線の差分の絶対値を最小化するルール」の最小絶対値法という手法もあります。

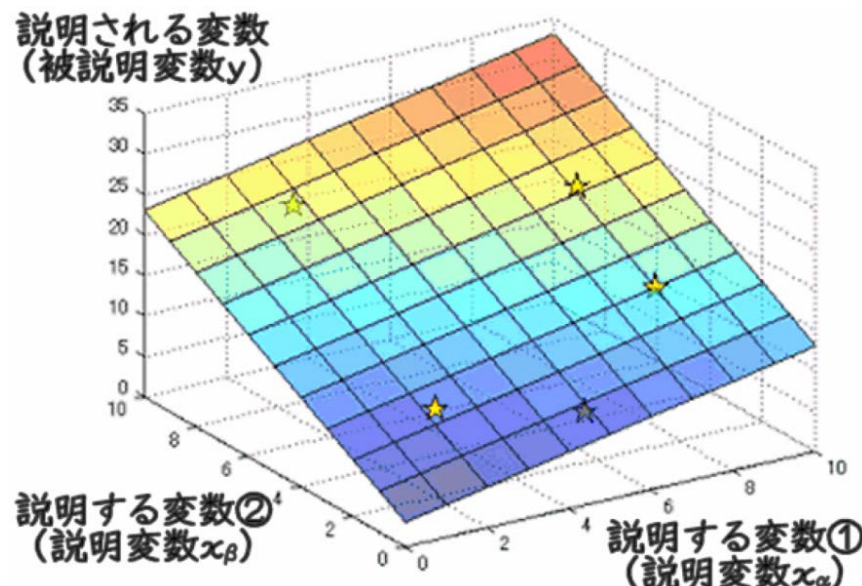


# 重回帰分析（説明変数が2種類以上の回帰分析）

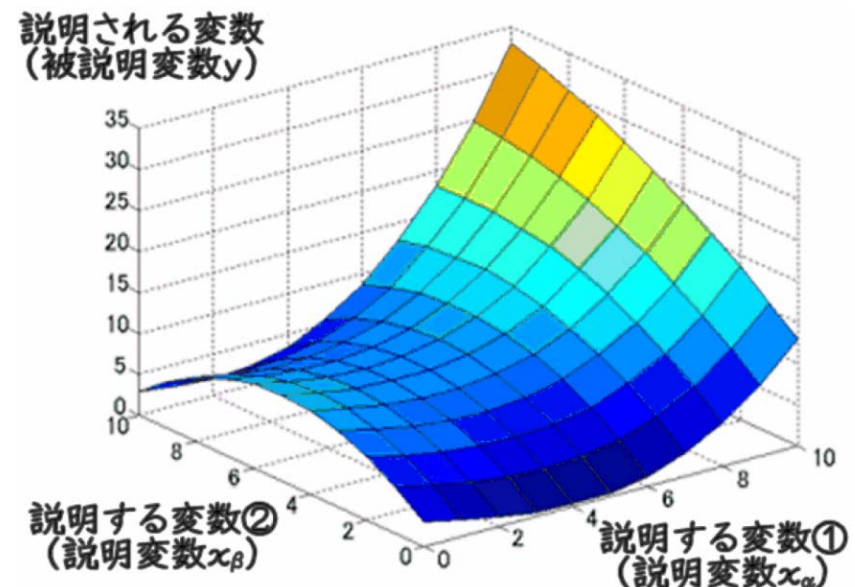
## ◆2種類以上の説明変数による回帰分析のことを重回帰分析と呼びます。

- Excelの分析ツールを利用すれば、「説明変数が2種類以上の重回帰分析」を行えます。
  - Excelの散布図における「近似曲線」では単回帰分析しかできませんが、Excelの分析ツールでは説明変数16種類までの重回帰分析を実行できます。
- 単回帰分析においては、最小化の対象は「星と直線の差分の二乗和」でしたが、重回帰分析においては当てはめる直線（⇒回帰線）が左下図のように板（⇒回帰平面）になります。
  - 説明変数が3種類以上で四次元以上になると想像しにくいですが、考え方は二次元から三次元に増える場合と同じです。
  - 回帰平面は各説明変数の平均値と被説明変数の平均値を通り、回帰平面の傾きは各説明変数から被説明変数への平均的な効果に対応します。
- 標準的な回帰分析では星と直線・平面との差分を考えますが、（直線・平面を前提としても）説明変数をあらかじめ2乗するなどの変換しておき、説明変数を元の尺度に戻せば、右下図のように曲線（曲面）の関係を表すこともできます。
- 「xの2乗」を説明変数に入れると、xが1、2、3と増えた場合に「xの2乗」は1、3、9と増加し、もとのxの尺度では曲線（曲面）の回帰線となります。

重回帰分析のイメージ



説明変数を変換して曲線の関係も分析可能





# 回帰分析用のデータセット（ダミー変数の作成）

## ◆土日祝を表すダミー変数を追加して、回帰分析用のデータセットを作成します。

- [東京店：飲料販売量(本)] を被説明変数として [東京：平均気温(℃)] と [東京：平均湿度(%)] を説明変数として、重回帰分析を行うことを考えます。
  - シート[3]の相関行列で示したようにシート[1]の全ての変数は正の相関関係にあります。相関関係としては、気温または湿度が高まれば、飲料販売量が増える直線的な関係にあります。しかし、相関係数や単回帰分析では気温か湿度のどちらかだけの効果か、両方ともに効果があるのかを判別できません。

 シート[1]から [日付] [曜日 (休日)] [東京：平均気温(℃)] [東京：平均湿度(%)] の各変数をコピーして、新規シートのA列～D列へ貼りつけます。

- 飲料販売量に影響を与える説明変数として、「平日」か「土日祝」の違いをダミー変数として追加することを考えます。

 E2のセルに『=IF(OR(B2="月",B2="火",B2="水",B2="木",B2="金"),0,1)』と入力すると、1が表示されます。

- Excel関数IFの中にExcel関数ORを入れ込み、B2の値が「月」「火」「水」「木」「金」のいずれかと一致すれば0を出力し、そうでなければ1を出力します。

 表示後にE2のセルの右下の■をダブルクリックして、0か1の表示を最下段まで引き延ばし、変数名に「土日祝ダミー」と入力します。

- 作成した [土日祝ダミー] は、[曜日 (休日)] に基づき、平日であれば0、土日祝であれば1の値を持ちます。
- 本講座の冒頭で「回帰分析は連続的な値（間隔尺度や比率尺度）をとる変数間の分析」と紹介しましたが、説明変数に関しては、名義尺度や順序尺度も「**0か1の値をとるダミー変数**」として、回帰分析に利用できます。
  - より高度な回帰分析には、被説明変数を名義尺度や順序尺度とするものもありますが、標準的な回帰分析の範疇外となります。

 シート[1]から [東京店：飲料販売量(本)] をコピーして、新規シートのF列へ貼りつけます。

作成した回帰分析用のデータセット

	A	B	C	D	E	F
1	日付	曜日 (休日)	東京：平均気温(℃)	東京：平均湿度(%)	土日祝ダミー	東京店：飲料販売量(本)
2	2016/1/1	金<休>	7.5	45	1	296
3	2016/1/2	土<休>	7.3	51	1	414
4	2016/1/3	日<休>	9.3	61	1	343
5	2016/1/4	月	9.2	60	0	514
6	2016/1/5	火	10.9	51	0	429

# Excelの「分析ツール」の回帰分析における指定

## ◆回帰分析のダイアログボックスに、被説明変数と説明変数を指定します。

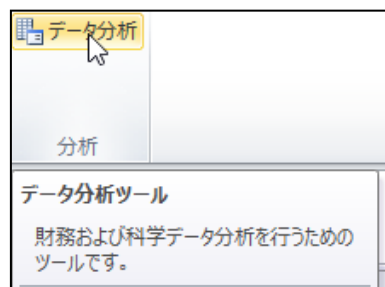
- Excelの「データ分析」のダイアログボックスにおいて、「回帰分析」を指定し、入出力の変数と出力先を指定します。

📄 Excel画面上側の「データ」タブの「データ分析」をクリックし、ダイアログボックスから「回帰分析」を選択し、「OK」をクリックします。

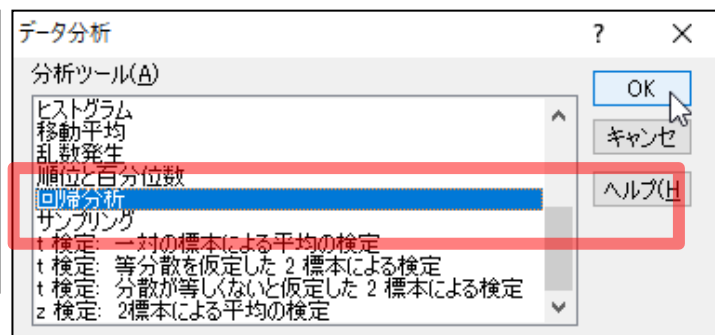
📄 入力Y範囲には、「説明される変数（1列のみ）」として「F列の範囲」を指定し、入力X範囲には「説明する変数（複数列でも可）」として、「C列～E列」を指定します。ボックス内の「ラベル」にチェックを入れて「OK」をクリックします。

- 回帰分析の結果を異なるExcelシートに出力したい場合は初期設定の「新規ワークシート」にチェックを入れて「OK」をクリックします。一方で、分析用のデータセットと同じシート内に表示したい場合は、「一覧の出力先」に分析結果を表示したい範囲の左上のセルを指定します。

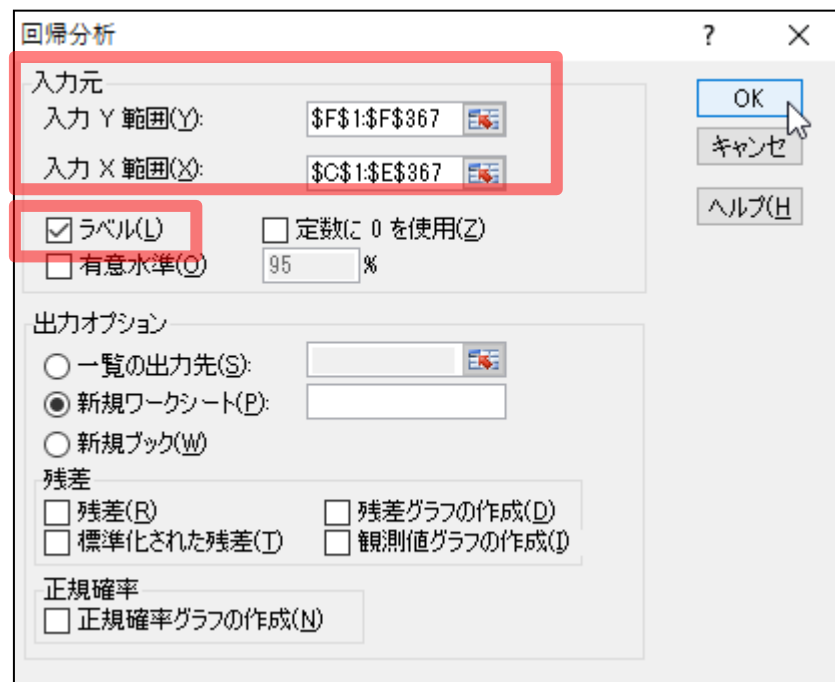
「データ分析」をクリック



「回帰分析」を選択して「OK」をクリック



「入力Y範囲」「入力X範囲」を記入して「OK」をクリック



# Excelの回帰分析における結果出力

## ◆Excelの分析結果は、「回帰分析全体」「個別の説明変数」の2種に分かれて表示されます

- Excelの回帰分析の結果出力は3ブロックに分かれ、上側の2つのブロックが【回帰分析全体に関する出力】に相当し、最下段のブロックが【個別の説明変数に関する出力】の結果になります。

Excelの回帰分析の結果出力

概要									
回帰統計									
重相関 R	0.720								
重決定 R <sup>2</sup>	0.518								
補正 R <sup>2</sup>	0.514								
標準誤差	62.689								
観測数	366								
分散分析表									
	自由度	変動	分散	観測された分散比	有意 F				
回帰	3	1530902	510300.6	129.849	4.16E-57				
残差	362	1422641	3929.946						
合計	365	2953542							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613	
東京: 平均気温(°C)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847	
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946	
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037	

回帰分析全体に関する出力

結果出力の上段

結果出力の中段

個別の説明変数に関する出力

結果出力の下段

- この講座では回帰分析の結果出力の上段と中段を簡潔に説明し、予測等に利用する下段を重点的に説明します。

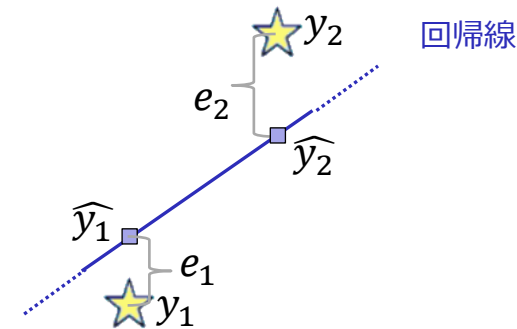
# 出力の上段における決定係数〈重決定 R2〉

## ◆上段出力の2行目の決定係数〈重決定 R2〉は、回帰分析の当てはまりの指標です。

- 上段出力の2行目の**決定係数〈重決定 R2〉**は、**0から1の値をとる回帰分析の当てはまりの指標**です。
  - Excelの回帰分析の出力にある〈重決定 R2〉〈重相関 R〉〈補正 R2〉という用語は統計学においても、実際の分析レポートにおいても、ほとんど使われず、Excelの分析ツールの独自用語です。本教材では、Excelの独自用語は〈〉で括って示し、統計学で一般的に使われる用語の後に記載しています。
  - 散布図における単回帰分析でもR2として表示され、重要度の高い2行目の決定係数〈重決定 R2〉から説明した後、1行目の〈重相関 R〉を説明します。

- 回帰分析では被説明変数を変数  $y$  で表す時、回帰線による**予測値**を $\hat{y}$ で表し、**推定エラー**を  $e$ で表し、 $i$  番目の標本に関して  $y_i = \hat{y}_i + e_i$  と書けます。

予測値と推定エラーのイメージ



- 予測値の  $\hat{y}$  は「ワイハット」と読み、推定エラー (error) の頭文字に由来する  $e$  は「イー」と読みます。
  - 予測値は当てはめ値と呼ばれることもあり、推定エラーは残差と呼ばれることもあります。
- 予測値  $\hat{y}$  の平均値は、被説明変数  $y$  の平均値  $\bar{y}$  に一致する性質があり、
 
$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2$$
 という式が必ず成り立ちます。
  - この式は「被説明変数の偏差平方和=予測値の偏差平方和+推定エラーの二乗和」を意味しています。
  - 推定エラー  $e$  の平均値は0という性質があり、「推定エラーの二乗和」は「推定エラーの偏差平方和」でもあります。

- この式の両辺を左辺で割って移項すると、**決定係数〈重決定 R2〉** =  $\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$  となり、

**決定係数〈重決定 R2〉**は  $\frac{\text{予測値の偏差平方和}}{\text{被説明変数の偏差平方和}} = 1 - \frac{\text{推定エラーの二乗和}}{\text{被説明変数の偏差平方和}}$  を意味しています。

- 上記の式の各項は、二乗した値の総和をとっているため0以上の値となり、決定係数は0以上1以下の値になります。
- **決定係数〈重決定 R2〉は、被説明変数の偏差平方和に占める予測値の偏差平方和を示し、被説明変数の変動を予測値の変動で説明できる割合に相当します。**

# 回帰分析の結果出力の上段

## ◆結果出力の上段は、回帰分析全体に関する出力を示しています。

- 出力2行目の**決定係数**〈重決定 R2〉は、前スライドで示した**回帰分析の当てはまりの指標**です。
  - 今回の出力では、予測値の変動によって被説明変数の変動の51.8%を説明できていることを示しています。

- 出力1行目の〈**重相関 R**〉は、2行目の決定係数〈重決定 R2〉の正の平方根です。
  - 〈重相関 R〉は、Excelの独自用語であることに加えて、2行目の決定係数〈重決定 R2〉から算出できることもあって、実際の分析資料において〈重相関 R〉を表示することは、ほとんどありません。

- 出力3行目の**自由度調整済み決定係数**〈補正 R2〉は、**自由度（＝標本数-説明変数の種類数）**を考慮した回帰分析の当てはまりの指標です。

$$\text{自由度調整済み決定係数} \langle \text{補正 R2} \rangle = 1 - (1 - \text{決定係数}) \cdot \frac{(\text{標本数} - 1)}{(\text{標本数} - \text{説明変数の種類数})}$$

自由度

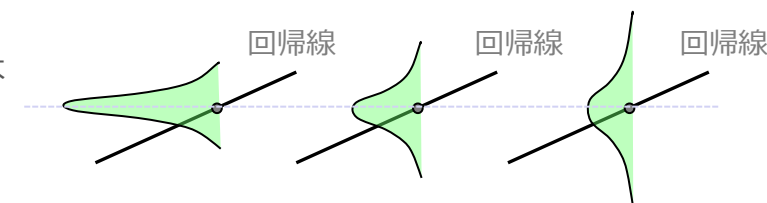
結果出力の上段

概要	
回帰統計	
重相関 R	0.720
重決定 R2	0.518
補正 R2	0.514
標準誤差	62.689
観測数	366

- 決定係数は前スライドに示したように被説明変数の変動に占める（全ての説明変数での）予測値によって説明できる割合を示しています。無関係な説明変数であっても、説明変数の種類を増やせば、予測値で説明できる割合に相当する決定係数は必ず上昇します。このため、説明変数を増やすことによって低下する要素として、上記の式の分数箇所を加えた当てはまりの指標が「自由度調整済み決定係数」です。
- 標本数が膨大であれば、上記の式の分数箇所は1に接近し、決定係数〈重決定 R2〉と自由度調整済み決定係数〈補正 R2〉は、僅差となります。
- 自由度の考え方は、講座3-2で示した「（母集団として）の分散と標本分散」の関係と同一であり、標本数と同じだけ説明変数の種類があれば、全ての標本を説明できて当たり前なので「説明できない可能性がある余地」に相当する自由度（＝標本数-説明変数の種類数）で調整しています。
- 出力4行目の〔**推定エラーの**〕標準誤差は、**回帰式の推定エラー（e）の平均的なバラツキ**を示しています。

右側ほど〔推定エラーの〕標準誤差が大きい

- 回帰分析の出力下段にも「標準誤差」と同様の表記があるなど、Excelの出力において、語句の混同を招く部分、言葉足らずの部分は本教材において〔〕で補足します。
- 決定係数の導出にも利用した「推定エラーの二乗和」を自由度で割ると、〔推定エラーの〕不偏分散が導出でき、その正の平方根をとることで〔推定エラーの〕標準誤差となります。
- 変数の分布に関するバラツキを標準偏差、推定値に関するバラツキを標準誤差といいます。





# 回帰分析の結果出力の中段（帰無仮説の考え方）

## ◆結果出力の中段は、回帰分析の全ての変数が無意味な場合の検定結果を示しています。

- 回帰分析の結果の中段にある〈分散分析表〉は、『切片以外の全ての説明変数は無効⇒切片以外の説明変数の真の係数は全て0である』という帰無仮説の検定を行っています。
- 利用可能な標本を用いて変数間の関係を推測する場合、誤差等の影響によって考察すべき母集団の変数間の関係を正しく把握できません。「説明変数は無効」「真の係数はゼロ」とは、考察すべき母集団において「変数間が無関係であること」を指しています。

### 帰無仮説の考え方

統計学では『効果がない（無効）』や『異なっていない（同質）』を主張する仮説を**帰無仮説（きむかせつ）**といいます。一方、帰無仮説の反対側の『効果がある（有効）』や『異なっている（異質）』を主張する仮説を**対立仮説**といいます。対立仮説を直接肯定することが難しいため、帰無仮説を否定することで**間接的に対立仮説を肯定するという手続き**をとります。なお、統計学の用語では「特定の判定基準のもとで帰無仮説を否定する」ことを「**帰無仮説を棄却（ききやく）する**」や「**統計的に有意（ゆうい）な効果/違いがある**」といいます。

- 中段の一番右側にある〈有意F〉は『切片以外の説明変数の真の係数は全て0である』という帰無仮説のもとで、偶然の誤差の影響によって標本の関係が観測されてしまう確率の上限を示しています。
- 〈有意 F〉はExcelの独自用語です。統計学では、この指標の利用頻度は少ないものの『複数制約のP値』と呼ぶ方が一般的です。

### 結果出力の中段

分散分析表					
	自由度	変動	分散	観測された分散比	有意 F
回帰	3	1530902	510300.6	129.849	4.16E-57
残差	362	1422641	3929.946		
合計	365	2953542			

「4.16E-57」は「 $4.16 \times 0.1^{57}$ 」を表し、0が57個並ぶほど、限りなく0に近い数値です。  
『全ての変数が無効』であることを前提とすれば、極めて不自然なことが起こったことを意味しています。

- 今回の出力のように〈有意F〉が十分にゼロに近ければ、「『**全ての変数が無効**』という帰無仮説を前提とすれば、**確率的に極めて珍しいことが起こった**、つまりは『**前提としていた帰無仮説は不自然で、ほぼ確実に効果のある説明変数がある**。』という解釈になります。
- 統計学は確定的な結論が出る学問ではありませんが、ある仮定を前提にして「無理がある/不自然である」ことを示して、前提としていた仮定を吟味する帰無仮説の検定の考え方は、数学における背理法と類似しています。



# 〔推定〕係数

◆ 〔推定〕係数は「説明変数1単位の増加→被説明変数への効果」の推定値を指しています。

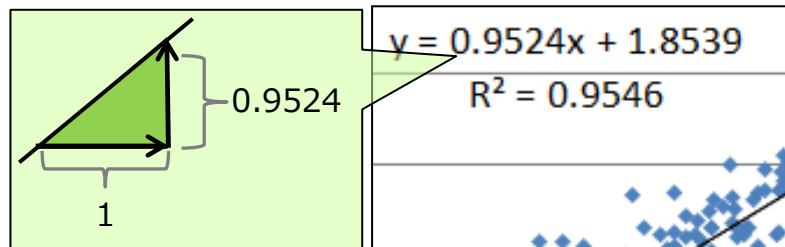
- 切片の〔推定〕係数は、他の全ての説明変数がゼロの場合における回帰分析による予測値に該当します。
  - 今回の出力では、仮想的に気温0℃、湿度0%の状態を想定した場合、平日の飲料販売量の予測値が約375本であることを示しています。

結果出力の下段（〔推定〕係数）

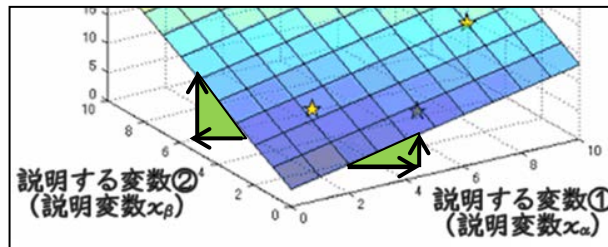
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

- 切片以外の〔推定〕係数は、説明変数の1単位の増加→被説明変数への効果の推定値を示しています。
  - 今回の出力では、気温が1℃上がれば飲料販売量は約6本増え、湿度が1%上がれば、飲料販売量が0.5本増えることを示しています。
  - 効果の方向の仮定を誤って逆にすると、「飲料販売量が上がれば、気温が上がる」ので「涼しくするために飲み物を控えよう」という誤った結論になります。
  - 説明変数の単位を1℃単位から10℃単位にするなど、入力値を0.1倍にした場合、推定係数は調整されて10倍になります。
- ダミー変数に関する〔推定〕係数は、ダミー変数が1となる場合に予測値が変化する水準を示しています。
  - 今回の出力では、土日祝ダミーの推定値は「-87」であり、平日を基準として土日祝の回帰線は、87本少なくなるように平行移動します。
  - 今回のダミー変数は、利用頻度の高い「定数項ダミー」という回帰線が平行移動するタイプで設定しましたが、「係数ダミー」と呼ばれる他の説明変数の推定値（傾き）を変えるタイプもあります。[例：平日と土日祝では「気温→飲料販売量」の効果の大きさ（傾き）が異なる想定]

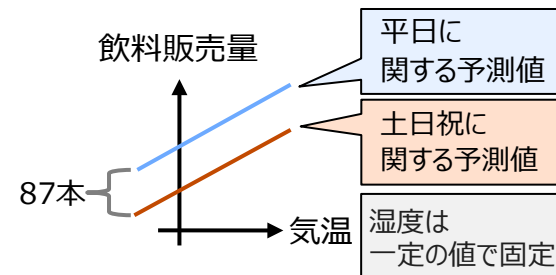
単回帰分析における推定係数（傾き）



重回帰分析における推定係数（傾き）



ダミー変数の出力結果イメージ



# 〔推定係数の〕標準誤差と〔信頼区間の〕下限・上限

◆〔推定係数の〕標準誤差と〔信頼区間の〕下限・上限は、推定係数の不確かさに関する指標です。

- 〔推定〕係数は利用可能な標本から推定するため、母集団において真の関係を表す係数からの乖離が見込まれます。
  - ・ 真の係数とは、無限個の標本に対応する母集団（考察の対象とする全ての標本）を利用した場合に確認できる正しい関係を指しています。
  - ・ 乱数に基づくシミュレーションにおいては真の係数を設定できますが、実証分析における真の係数は分析者にとって最後まで分からない数値です。
- Excel出力の〔推定係数の〕標準誤差は、「推定係数の不確かさ（⇒推定精度の悪さ）」に該当します。
  - ・ 説明変数の単位を1℃単位から10℃単位にするなど、入力値を0.1倍にした場合、〔推定係数の〕標準誤差は調整されて10倍になります。
- 〔信頼区間の〕下限および上限の95%は、信頼係数95%で真の係数がありそうな範囲を示しています。
  - ・ 回帰分析実行時のダイアログボックスで90%や99%の信頼係数を入力して指定することができ、結果出力における右側の2列の値が変化します。
  - ・ 〔推定〕係数によって一点で示されている値が、真の係数に一致すると考えるのは楽観的ですが、真の係数は推定係数を中心にその付近に存在すると考えるのが自然です。

## 信頼係数

ここで**確率**という言葉を使わず、**信頼係数**という言葉を用いるのは「**真の係数は分析者にはっきりと分からないだけで、確定した値がある**」という考え方に由来します。例えば、地震の研究者への『3世紀の日本で震度7以上の大地震が起こったか？』という問いに対して、「私は十中八九、起こったと考える」とは答えられても、「確率80%～90%で起こった」とは答えません。ただし、信頼係数という言葉に馴染めなければ、**信頼係数は「（利用可能な標本に基づく分析者の）主観的な確率」と読みかえてもかまいません。**

- 〔信頼区間の〕下限および上限 95%の値は、おおよそ推定係数 $+(±2)×〔推定係数の〕標準誤差$ となります。
  - ・ 今回の出力では、気温の推定係数の信頼区間95%の上限の概数は $5.88+2×0.491=6.862$ （Excel出力：6.847）であり、下限の概数は $5.88-2×0.491=4.898$ （Excel出力：4.914）であり、それぞれ近似値になっています。

結果出力の下段（〔推定係数の〕標準誤差、〔信頼区間の〕下限・上限）

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

# 【参考】乱数シミュレーションに基づく回帰分析

## ◆推定係数の標準誤差は、変動する要因を乱数シミュレーション等で確認します。

- 推定係数の標準誤差は、下の表内の要因によって減少します。

推定係数の標準誤差を低下させる要因と確認方法

項目	〔推定係数の〕標準誤差を低下させる要因	〔推定係数の〕標準誤差の増減との関係の説明	〔推定係数の〕標準誤差が低下することの確認方法
[1]	標本数の増加	標本数が多ければ、推定精度は高まるため、推定係数の標準誤差は小さくなります。	シート[4]のデータセットを縦に並べて2年分（732標本）の扱いで回帰分析を実行し、比較してください。
[2]	各説明変数のバラツキ（標準偏差）の増加	各説明変数のばらつきが大きい方が、推定係数を測りやすく、推定係数の標準誤差は小さくなります。	シート[参2]における説明変数の標準偏差を指定するF4およびF5のセルの値を増加させてください。
[3]	説明変数同士の相関係数の絶対値の低下	説明変数同士が相関が小さいと、どの説明変数の効果かを判別しやすく、推定係数の標準誤差は小さくなります。	シート[参2]における説明変数間の相関係数を指定するB6のセルの値を0に近づけてください。
[4]	回帰式全体に関するエラーのバラツキ（標準偏差）の低下	回帰式全体に関するエラーのバラツキが小さいと、推定係数を測りやすく、推定係数の標準誤差は小さくなります。	シート[参2]における回帰式のエラーの標準偏差を指定するB7のセルの値を減少させてください。

- 項目[1]に関して、膨大な標本数（→標本数の多いビッグデータ）では、推定係数の標準誤差は0に接近し、ほとんど全ての帰無仮説を棄却できます。
- 項目[3]に関して、説明変数同士の相関係数が±1（→完全な多重共線性）の場合、どちらの説明変数の効果かを判別できず、分析できなくなります。

## □ シート[参2]では、シミュレーションに基づいて回帰分析の出力の挙動を確認できます。

シート[参2]乱数シミュレーションに基づく回帰分析

キーボードの[F9]または「数式」タブの「再計算実行」をクリックして、出力の挙動を確認してください。

水色のセルを変化させて、設定を変更できます。キーボードの[F9]ボタンまたは「数式」タブの「再計算実行」で乱数が再発生し、回帰分析の出力が再計算されます。					
切片	15	標本数	50		
$X_\alpha$ の係数	3	$X_\alpha$ の平均	10	$X_\alpha$ の標準偏差	7
$X_\beta$ の係数	5	$X_\beta$ の平均	20	$X_\beta$ の標準偏差	3
$X_\alpha$ と $X_\beta$ の相関係数	0.3	←項目[3]の確認は、この値を0に近づける[範囲:-1~1]			説明変数 $x_\alpha$ と $x_\beta$ は二次元正規分布にて発生
回帰式のエラーの標準偏差	9	←項目[4]の確認は、この値を減少させる[範囲:0~]			回帰式のエラーは、正規分布にて発生
真のデータ発生プロセス:	$Y=15+3X_\alpha+5X_\beta+N(0, 9^2)$				
回帰分析の出力:	$Y=15.15+3.41X_\alpha+4.77X_\beta+N(0, 7.1^2)$				
					↓ Excelの分析ツール風の出力

# t[値]

## ◆t[値]は、基準精度で評価した推定係数に該当します。

$$t\text{値} = \frac{\text{推定係数}}{\text{推定係数の標準誤差}}$$

### ● t値は、推定係数の標準誤差で割ることで基準精度で評価した推定係数です。

- 今回の出力においては、いずれの説明変数のt値も絶対値で2を超えており、説明変数のそれぞれに被説明変数への効果があったと見なせます。
- 説明変数の単位を1℃単位から10℃単位にするなど、入力値を0.1倍にした場合、分子と分母のそれぞれが10倍されてt値は変わりません。

結果出力の下段 (t[値])

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(℃)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

### ● t値を用いて『説明変数の被説明変数への真の効果が0である』という帰無仮説を検定することができます。

#### t分布と標準正規分布

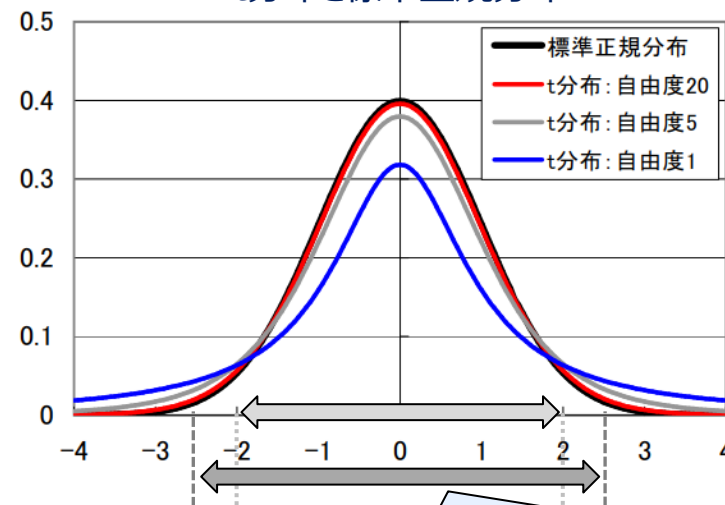
説明変数の「真の効果がゼロ」という帰無仮説が正しければ、**t値はt分布**に従うことが知られています。t分布は、右図のように0を中心に左右対称にばらつく山形の分布です。なお、**t分布は自由度（標本数-説明変数の種類数）が上がると、標準正規分布に接近します**。自由度が20以上であれば、実務上は標準正規分布と考えて差し支えありません。

自由度の高いt分布や標準正規分布において、絶対値で2からはみ出す面積の割合は、プラス側とマイナス側のそれぞれ約2.5%で両側約5%です。

実際の分析で利用頻度の高い「両側5%の基準での検定」の考え方

t値の絶対値が2未満	⇒帰無仮説の前提で5%以上は発生する自然なt値 ⇒帰無仮説が誤っているとは考えにくい。
t値の絶対値が2以上	⇒帰無仮説の前提で5%以下でしか発生しない不自然なt値 ⇒帰無仮説が誤っていると考え、効果のある変数と見なせる。

t分布と標準正規分布



帰無仮説を前提とすれば、  
 両側5%基準：95%が±2.0の範囲に収まる ⇔  
 両側1%基準：99%が±2.6の範囲に収まる ⇔



# P-値

## ◆P-値は、統計的検定の境目となる確率を表しています。

- 前スライドに示したt値は、次のような場合に利用しにくい指標となります。
  - ◆ 両側5%以外の基準で検定を行う場合 (⇒ t値の絶対値の2を基準とする目安が利用できない)
  - ◆ 自由度 (標本数-説明変数の種類数) が20未満などの場合 (⇒ t分布の標準正規分布への接近が不十分)
- こうした場合に便利で、**自由度を考慮した統計的検定の境目を出力しているのがP-値**です。

(自由度の高い) 回帰分析において、t値が2以上2.6未満の範囲に入った場合は「両側5%基準では帰無仮説が棄却できる (⇒両側5%で有意)」の一方で「両側1%基準では帰無仮説が棄却できない (⇒両側1%では有意ではない)」ことになります。

➡ 1%と5%の間に境目となる基準の値があったはずで、P-値はこの境目となる値を導出しています。

- 今回の出力では、湿度のt値は2.115であり、5%と1%の間の検定の境目があるはずで、右隣のP-値を見ると、それが3.5%であることが分かります。

結果出力の下段 (P-値)

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	374.963	14.569	25.738	8.80898E-84	346.314	403.613	346.314	403.613
東京: 平均気温(°C)	5.880	0.491	11.965	4.98243E-28	4.914	6.847	4.914	6.847
東京: 平均湿度(%)	0.490	0.232	2.115	0.035070468	0.035	0.946	0.035	0.946
土日祝ダミー	-86.739	6.968	-12.449	7.40285E-30	-100.441	-73.037	-100.441	-73.037

- 両側5%の判定基準に関しても、「t値が絶対値で2を超えているか？」の目安よりも「P-値が0.05を切っているか？」の方が、自由度を考慮していて正確です。

### 統計的検定と説明変数の効果

t値やP-値を用いた検定は『推定精度をふまえて、効果が0なのか？』を基準に判定しています。このため、「『推定係数の絶対値が小さくとも、推定精度が高い変数』を効果がありそうな説明変数」と判定します。一方で「『推定係数の絶対値が大きくとも、推定精度が低い変数』を効果がなさそうな説明変数」と判定します。t値やP-値に基づく効果のある変数の表示は、良くも悪くも不確かさを考慮しているので、推定係数の（絶対値の）大きさに基づく表示や利用と使い分ける必要があります。

# 分析結果の考察と活用方法

## ◆ 回帰分析の分析結果は、説明変数の効果測定や予測に活用できます。

- 今回の出力では**決定係数**〈重決定  $R^2$ 〉が0.518であり、予測値の変動によって、被説明変数の変動の51.8%の説明ができています。
  - 決定係数の評価に絶対的な水準はないものの、回帰分析の設定に改善の余地が大きい決定係数の水準であるともいえます。
  - 説明変数を追加すれば決定係数は必ず上がりますが、分析目的によっては低い決定係数でも問題のない場合もあります。
- 回帰分析の式の形を変更したり、説明変数を追加したりして、**回帰分析の設定を改善することが可能です。**
  - 今回の回帰分析の設定では、気温が「0℃→1℃」でも「30℃→31℃」でも、飲料販売量に同一の効果があると仮定していました。「気温の2乗」や「気温の3乗」を説明変数に追加することで、気温の水準によって飲料販売量への効果が異なる可能性を考慮することができます。
  - 今回の回帰分析の設定では、気温と湿度は独立して、飲料販売量に効果があることを仮定していました。「気温×湿度」を説明変数に追加することで、気温と湿度の相乗効果の可能性を考慮することができます。
  - 今回の回帰分析の設定では、「土日祝ダミー」によって平日と土日祝では飲料販売量の水準が異なる可能性を考慮しましたが、土・日・祝のそれぞれで飲料販売量の水準が異なる可能性があります。土・日・祝のそれぞれに異なるダミー変数を設定することもできます。
  - 実証分析において、回帰分析の式の形はどうやっても近似にすぎませんが、現実を表すために「無理のない近似」となっている必要があります。
- 回帰分析では、**不確かさを踏まえて説明変数の被説明変数への効果を導出し、予測に利用**することができます。

### 回帰分析の活用例

- ◆ 「予想温度▲℃、予想湿度■%の平日における飲料販売量の予測」など、与えられた設定における予測が可能となる。

 **与えられた設定に基づく予測販売量を導出することで、最適な仕入れ量を検討することができる。**

- ◆ 複数の変数の関係を調整し、不確かさを考慮した上で、説明変数が被説明変数に与える効果を導出できる。

 **説明変数の効果の有無に関する水掛け論を防ぎ、統計的検定を踏まえて定量的・客観的な基準で評価・議論できる。**