

相関分析

三池 克明

ここでは Excel の機能の一つである「分析ツール」を使って相関分析をする方法を解説します。

データ分析に大変役立つ機能ですのでしっかりマスターしましょう。

—目 次—

1. 相関分析とは.....	1
1.1. 相関とは.....	1
1.2. 相関係数とは.....	1
1.3. 相関係数と相関の正負・強弱の関係	2
2. 表の作成	3
2.1. 入力	3
2.2. ウィンドウ枠の固定	3
3. 全ての組合せの相関係数を求める	5
3.1. 分析ツールを使えるようにする	5
3.2. 分析ツール「相関」を使って全ての組合せの相関係数を求める	6
4. 相関の強弱を分類する.....	9
5. 散布図やピボットテーブルを作成する	12

1. 相関分析とは

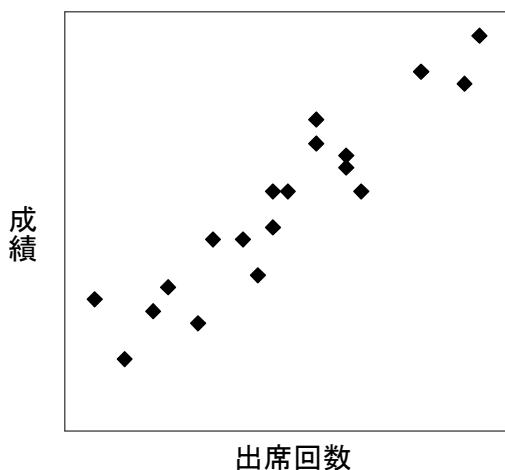
1.1. 相関とは

例えば、

- ある科目の出席回数と成績
- 身長と体重
- アルバイトに入った時間と給料

には何かしらの関係があることが分かります。

この「関係」の捉え方の一つが「相関」です。



このようにプロットした点
が一本の線に集まっているよ
うに見えることを「相関があ
る」という。

1.2. 相関係数とは

相関係数とは相関の度合いを示す数値です。

文字式では r であらわし、その値の範囲は

$$-1 \leq r \leq 1$$

と、なります。

※計算式は割愛します

1.3. 相関係数と相関の正負・強弱の関係

相関係数から、相関の正負と強弱を読み取ることができます。

- 正負

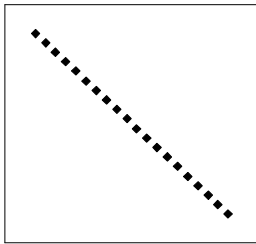
$r > 0$: 正の相関…点が右上がりの線に集まっているように見える

$r < 0$: 負の相関…点が右下がりの線に集まっているように見える

- 強弱

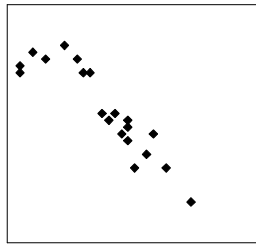
r が 1 または -1 に近い : 強い相関…点が線に集まっている

r が 0 に近い : 弱い相関…点が散らばっている

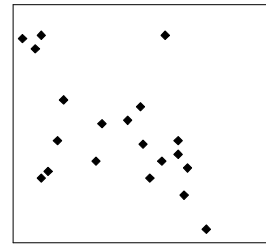


$r = -1$

←強い負の相関

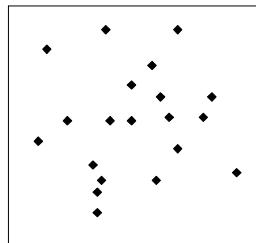


$-1 < r \leq -0.8$



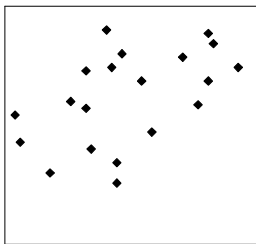
$-0.8 < r \leq -0.5$

弱い負の相関→



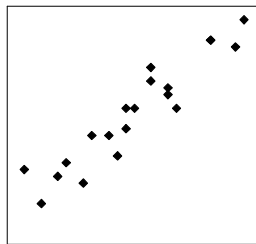
$-0.5 < r < 0.5$

相関が無い

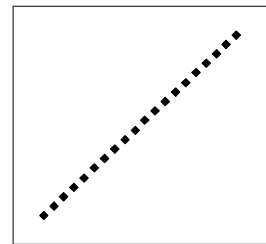


$0.5 \leq r < 0.8$

←弱い正の相関



$0.8 \leq r < 1$



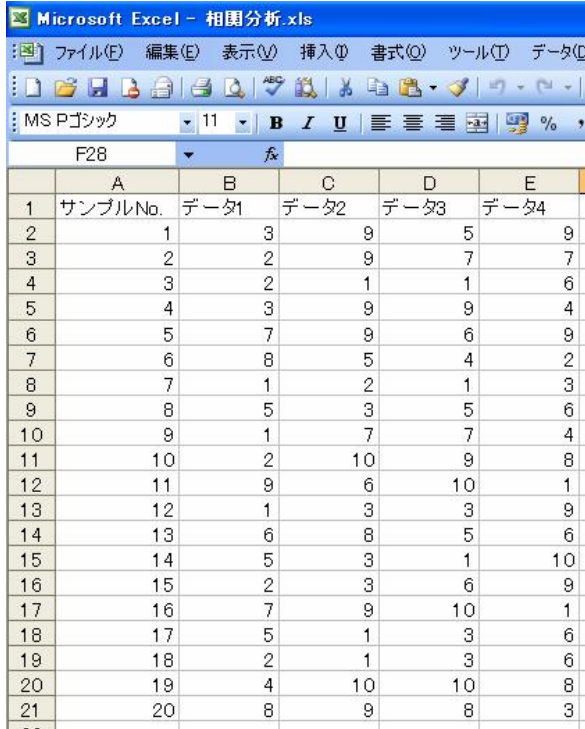
$r = 1$

強い正の相関→

(数値はあくまで目安です)

2. 表の作成

2.1. 入力



	A	B	C	D	E
1	サンプルNo.	データ1	データ2	データ3	データ4
2	1	3	9	5	9
3	2	2	9	7	7
4	3	2	1	1	6
5	4	3	9	9	4
6	5	7	9	6	9
7	6	8	5	4	2
8	7	1	2	1	3
9	8	5	3	5	6
10	9	1	7	7	4
11	10	2	10	9	8
12	11	9	6	10	1
13	12	1	3	3	9
14	13	6	8	5	6
15	14	5	3	1	10
16	15	2	3	6	9
17	16	7	9	10	1
18	17	5	1	3	6
19	18	2	1	3	6
20	19	4	10	10	8
21	20	8	9	8	3

左図のように表を作成します。
ファイル名は「相関分析.xls」
（“.xls”はExcelが自動的に付加
します）としましょう。

2.2. ウィンドウ枠の固定

セル B2 をクリックし、メニュー「ウィンドウ」－「ウィンドウ枠の固定」をク
リックします。



このようにアクティブセル（ここではセル B2）の左上を基準にウィンドウ枠の罫線が表示されます。

	A	B	C	D	E	F
1	サンプルNo.	データ1	データ2	データ3	データ4	
2	1	3	9	5	9	
3	2	2	9	7	7	
4	3	2	1	1	6	
5	4	3	9	9	4	

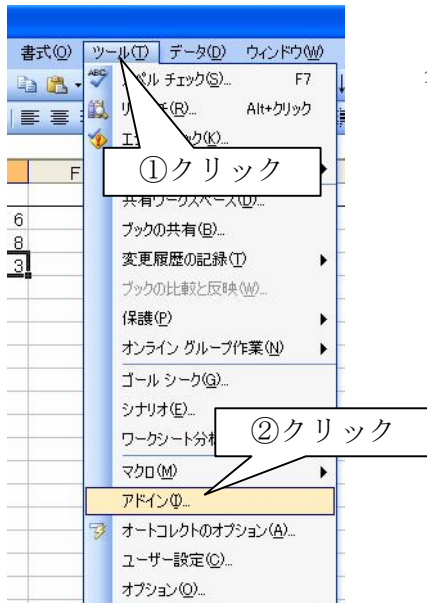
このようにワークシートをスクロールさせても A 列と 1 行目の表示が残ります。これにより、サンプルNo. やデータ項目名を見逃さずに作業をすることが可能になります。大きな表を扱うときに大変便利になりますので適宜活用しましょう。

	A	D	E	F	G	H
1	サンプルNo.	データ3	データ4			
19	18	3	6			
20	19	10	8			
21	20	8	3			
22						
23						

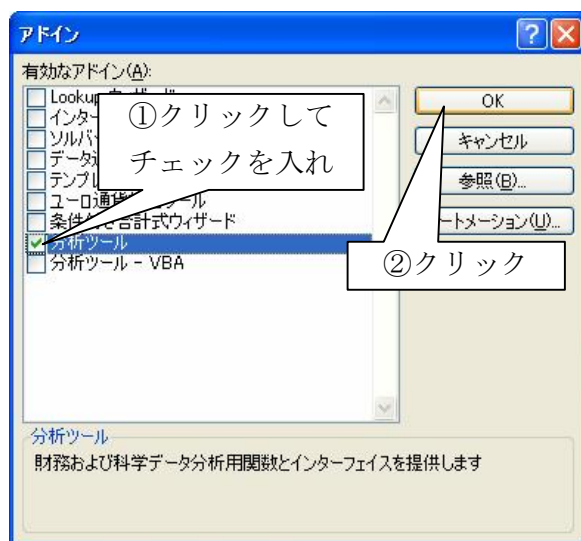
3. 全ての組合せの相関係数を求める

3.1. 分析ツールを使えるようにする

※メニュー「ツール」－「分析ツール」が無い場合は以下の作業を行ってください。



メニュー「ツール」－「アドイン」をクリックします。



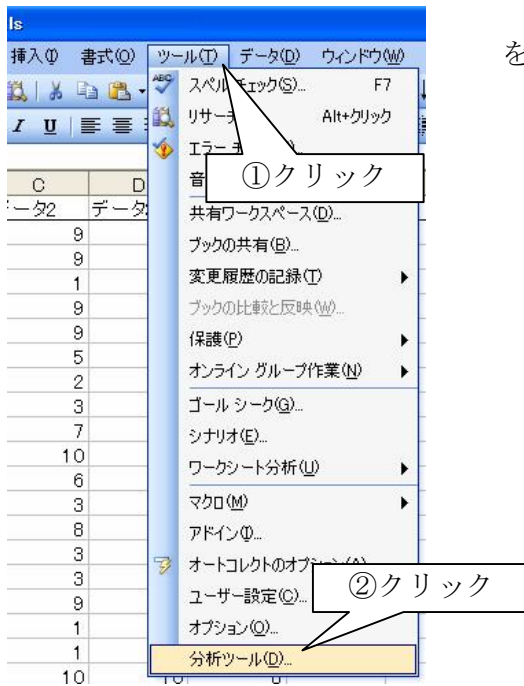
「有効なアドイン」欄の「分析ツール」をクリックしてチェックを入れ、「OK」ボタンをクリックします。

※ExcelのCD-ROMを求められることがあります

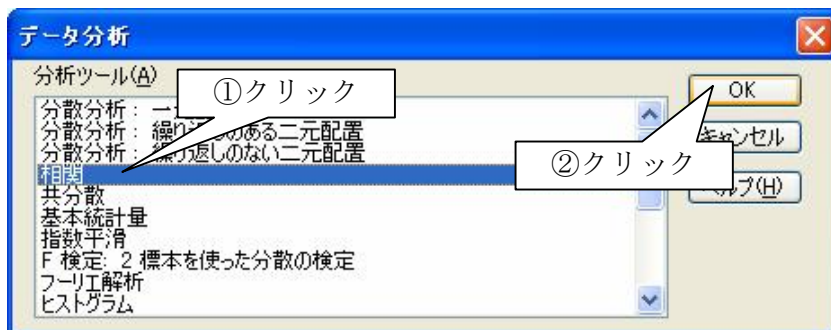
これで分析ツールが使えるようになります。

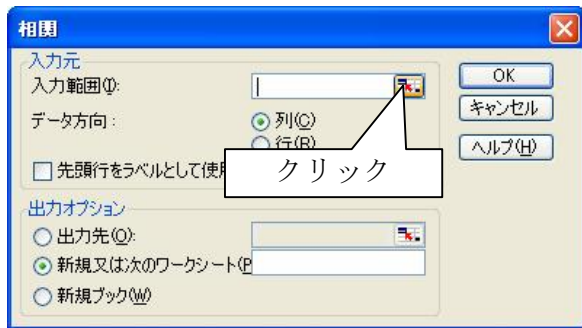
3.2. 分析ツール「相関」を使って全ての組合せの相関係数を求める


メニュー「ツール」－「分析ツール」をクリックします。




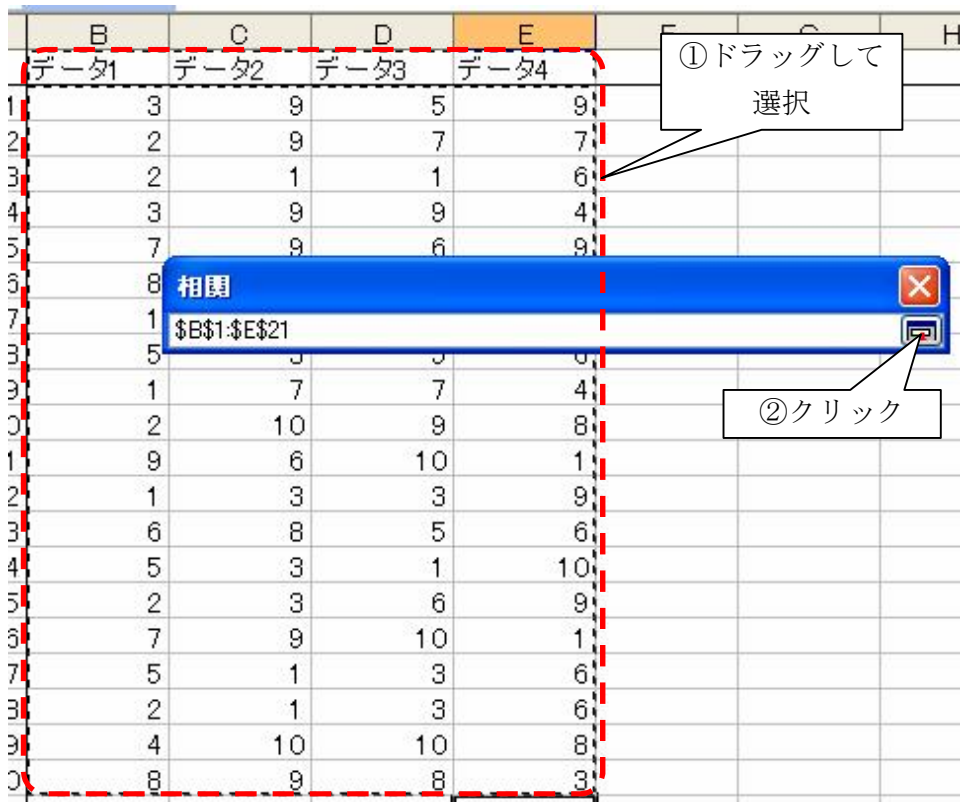
「分析ツール」欄の「相関」をクリックし、「OK」ボタンをクリックします。

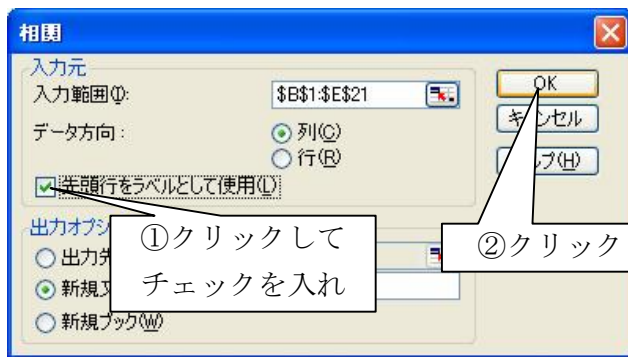




「相関」ダイアログボックスが表示されるので、「入力範囲」欄の  ボタンをクリックします。

セル B1～E21（データ 1～データ 4 の範囲）をドラッグして選択し、 ボタンをクリックします。





「先頭行をラベルとして使用」をクリックしてチェックを入れ、「OK」ボタンをクリックします。

	A	B	C	D	E
1		データ1	データ2	データ3	データ4
2	データ1	1			
3	データ2	0.257233	1		
4	データ3	0.32127	0.794522	1	
5	データ4	-0.43615	-0.05175	-0.2972	1

新規にワークシートが作成され、データ 1～データ 4 の全ての組合せの相関係数が表示されます。

データ 1 とデータ 4 の相関係数は
-0.43615
 であることが分かる

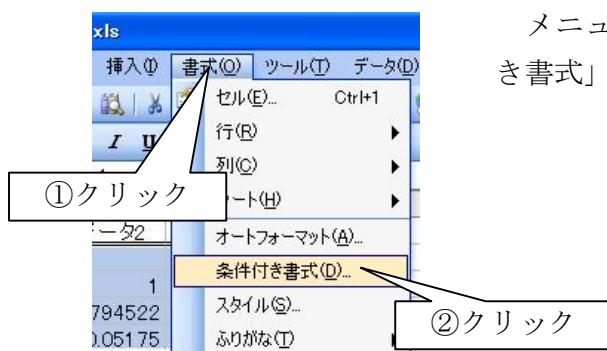
4. 相関の強弱を分類する

相関係数の表から相関の正負や強弱を見分けるのは困難です。
そこで「条件付き書式」をつかって色分けをします。

	A	B	C	D	E
1		データ1	データ2	データ3	データ4
2	データ1	1			
3	データ2	0.257233	1		
4	データ3	0.32127	0.794522	1	
5	データ4	-0.43615	-0.05175	-0.2972	1

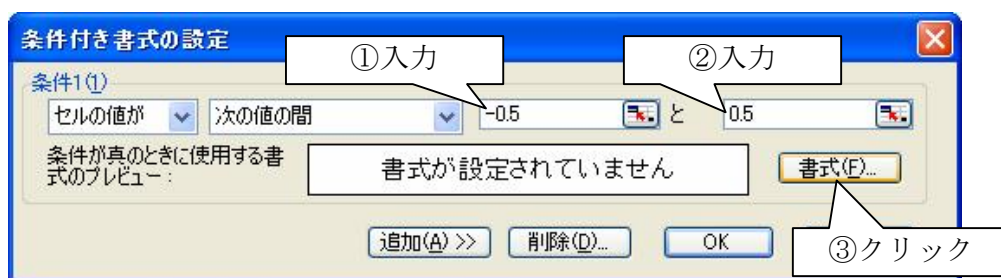
セル B2～E5（相関係数が書かれている範囲）をドラッグします。

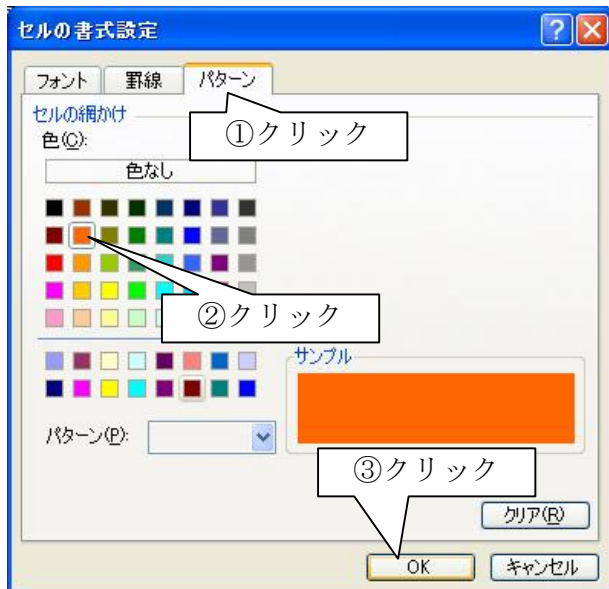
ドラッグして
選択



メニュー「書式」－「条件付き書式」をクリックします。

「条件付き書式の設定」ダイアログボックスが表示されるので下図のように値を入力し「書式」ボタンをクリックします。





「セルの書式設定」ダイアログボックスが表示されるので、条件を満たした場合の書式を設定します。
(ここではセルの色をオレンジにしています)

条件を追加するため「追加」ボタンをクリックします。



同じように条件と書式を設定します。
(ここでは値が-0.8 から 0.8 の間であればセルの色を黄色にするように設定しています)



設定が終わったら「OK」ボタンをクリックします。



相関係数の範囲に応じて色分けされました。

データ2とデータ3に正の相関が見られるようです。

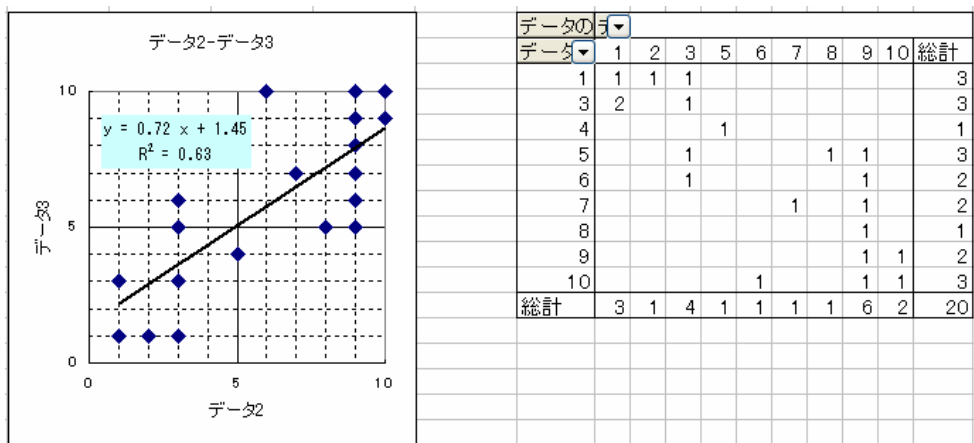
また、他は相関が無いことがわかります。

	A	B	C	D	E
1		データ1	データ2	データ3	データ4
2	データ1	1			
3	データ2	0.257233	1		
4	データ3	0.32127	0.794522	1	
5	データ4	-0.43615	-0.05175	-0.2972	1

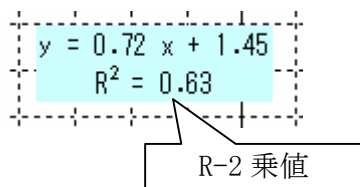
5. 散布図やピボットテーブルを作成する

※散布図やピボットテーブルの使い方は、別のテキストを参照してください。

データ2とデータ3で散布図とピボットテーブルを作りました。
見て分かるとおり、正の相関があることが分かります



散布図にある R^2 の式 ($R-2$ 乗値) は相関係数を2乗したものです。



$$0.794522^2 \doteq 0.630436$$

ですから間違いありませんよね。

参考文献

- (1) Microsoft Corporation : Microsoft Excel のヘルプ, Microsoft Excel
- (2) 縄田和満 : Excel による統計入門, 朝倉書店, 1996 年
- (3) 島 久代, 植木節子 : 統計初歩の初歩, 日科技連, 1999 年