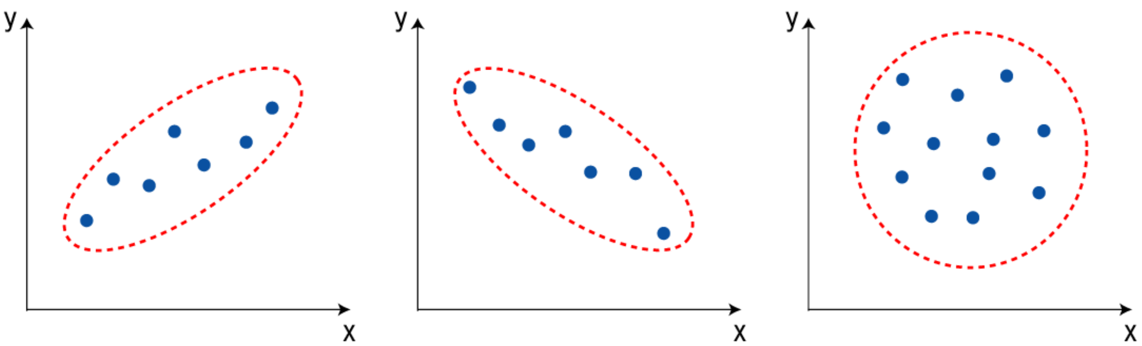
シミュレーション論Ⅱ　第13回資料　【データの分類と数値処理（2）】

【線形回帰の注意点】

- ■ 線形回帰では、2 変数の間に「直線的な」関係があることが前提となる
- ■ データ間の（直線的な）関係性の向きや程度を調べておく必要がある→相関分析

【相関分析】

- ■ 2 変数に関連性がある（片方が大きくなると、もう一方も大きくなる、など）場合、2 変数に「相関」があるという
- ■ **相関係数**：相関の度合いを表す数値で、Rで表す
- ■ 相関係数は－1～＋1の数値をとる
- －　片方が増えればもう片方も増える→相関係数は＋、正の相関
- －　片方が増えればもう片方が減る　→相関係数は－、負の相関
- ■ 絶対値が大きいほど、相関が強い（2 変数の間に強い関連性がある）



正の相関　　負の相関　　無相関

【相関係数の式】

n 組のデータ　$(x_1, y_1),\quad (x_2, y_2),\quad (x_3, y_3)\Lambda\quad (x_n, y_n)$　があるとき、相関係数 R は

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

ただし $\bar{x}, \bar{y}$　はそれぞれ x, y　の平均

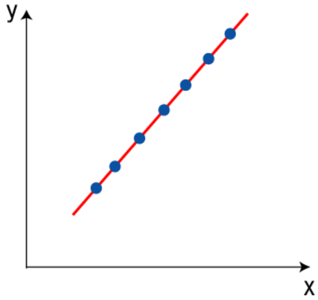
【相関係数の値と相関の度合い】

- ■ 相関係数の絶対値が　1　に近いほど相関の度合いは強くなる
- ■ 一般的には以下のような分類になる
- ■ 相関係数の絶対値が 1 の場合を「完全な相関」といい、データが一直線上に並ぶ場合に相当する

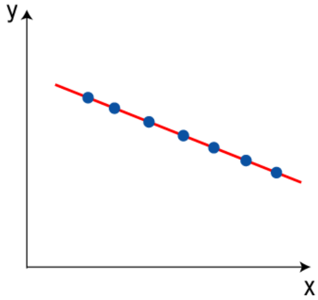
※相関の度合いの判定にはいくつかの基準があるが、いずれにしても絶対値が 1 に近いほど相関が強い（高い）

Rの値	相関の度合い
0	無相関
0<｜R｜≦0.2	ほとんど相関なし
0.2<｜R｜≦0.4	低い相関あり
0.4<｜R｜≦0.7	相関あり
0.7<｜R｜<1.0	高い相関あり
1.0 または-1.0	完全な相関

完全な正の相関
（R ＝ 1）



完全な負の相関
（R ＝ −1）



【練習：相関係数の計算】

以下のようなデータがあるとき、相関係数を計算してみよう。

$(x_1, y_1) = (1,19)$

$(x_2, y_2) = (2,23)$

$(x_3, y_3) = (3,35)$

$(x_4, y_4) = (4,37)$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

【相関係数の注意点】

- 相関係数の値はデータの個数やバラツキに大きく左右される
  →「たまたま」そうなったのか、関係性があるからそうなったのか
- 2 変数の間にちゃんと「相関があるかどうか」を調べてやる必要がある
  →無相関検定

【無相関検定】

- 2 変数の間に「相関がない」と仮定して（帰無仮説）、有意性の検定をおこなう
- 相関係数に意味があるかどうか（相関があるかどうか）
- 以下の表で、5%有意水準の値より相関係数 *R* が大きければ「有意」、つまり相関があるといえる（1%水準はさらに厳しい検定）

※実際には帰無仮説の棄却なので「無相関であるとはいえない」ということ

データ数（n）	5%有意水準	1%有意水準
3	0.997	1.000
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875
8	0.707	0.834
9	0.666	0.798
10	0.632	0.765
100	0.197	0.256
1000	0.062	0.081

【無相関検定の注意点】

- 相関係数の絶対値が 1 に近くても必ずしも相関があるとはいえない（無相関検定）
- 相関分析はあくまで「相関がありそう」という判定
  →まったく無関係のデータ間に強い相関が見られることもある
- ここまで述べてきた相関係数は実は「ピアソンの積率相関係数」というもので、2 変数間に線形の関係があることを前提にしたもの
- データが曲線に沿う場合や、他の関係がある場合には意味をなさない
-

【どうすれば適切な方法でデータの分類と整理ができるか？】

- 変数間の関係性を調べたいときには、まず散布図を描いてみる
- データの散らばりからある程度の仮説を立て、それに応じた分析方法を使う
- 線形の関係性が見られる場合→相関分析、線形回帰分析

【第13回のレポート】

■　気温とおでんの売り上げが以下の表のようになった。

（1）相関分析をおこない、気温とおでんの売り上げに相関があるかどうか、あるならばどのような相関か調べよ

- 相関係数を計算し、無相関検定により有意性を判定する
- 相関係数の値から相関の正負、強弱を調べる

（2）線形回帰をおこない、回帰直線の式を求めよ。

※無相関検定は *R* の絶対値で判定してよい

※電卓のない人は右の表も参考にしてください

気温（℃）	売り上げ（個）
18	
15	
10	
5	

参考：$\sqrt{\ }$の値

$\sqrt{111} = 10.535653$

$\sqrt{130} = 11.401753$

$\sqrt{1175} = 34.278273$

$\sqrt{7280} = 85.322916$

$\sqrt{12740} = 112.8716$

$\sqrt{13965} = 118.1736$