

◇ 教科書 p.44~p.56

● 2変数データ

- ・ = 各固体, 対象等に関し, 2種類の値を観測したデータのこと。
- ・ = 2つの変数の関係を調べることに適した, 縦軸と横軸の2つの項目で量を計測し, 分布を表現するために使うグラフ。

→ 2変数データの各変数の特徴を調べるには, 変数毎にグラフを作成したり, 代表値を計算すると良い。
 ただし, 各変数の特徴を調べるだけでは, 変数間の関係は分からない。
 (2変数データ (ペアの) 観測値をバラバラに扱ってはいけない。)

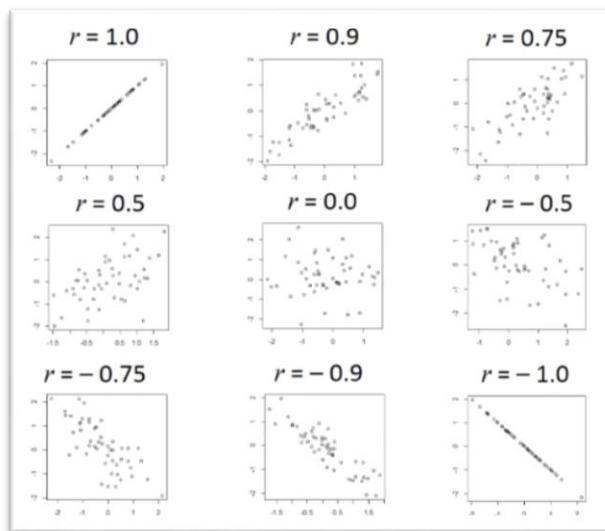
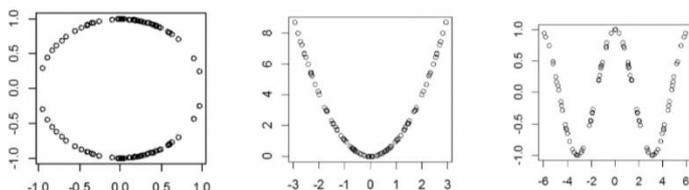
- ・ = ある特徴によって, 分割可能なデータのこと。
 ※ひとまとめに扱ってしまうと, 知りたい特徴が分からなくなる場合がある。
 層別に異なる傾向があるデータの散布図を作成すると, 以下のような現象が起こることがある。
 [1]本来は存在しない傾向が現れる。
 [2]本来存在する傾向が見えにくくなる。

● 相関係数

- ・ = 2つの直線的な関係の強さを示す指標。
 ◎相関係数の値は必ず-1以上1以下となる。
 ◎直線的な傾向が強いほど, 相関係数の値は-1または1の値に近づく。
 ◎一方の変数の増加とともに, もう一方の変数も増加する傾向にあれば, 相関係数の符号は正となり,
 一方の変数の増加とともに, もう一方の変数が減少する傾向にあれば, 相関係数の符号は負となる。

※データの数小さい場合は, 直線的な傾向の有無の議論はほとんどできない。
 ※外れ値は相関係数に影響を与えることがあるので注意が必要。

▽以下のようなデータは相関係数が0になる



相関係数が0に近い → 直線的な関係は無い。(2つの変数間に関係がないとはいきれない)

【課題】

エクセルの関数を用いて, 相関係数の演習をしてみよう。 相関係数…「correl」