

相関係数

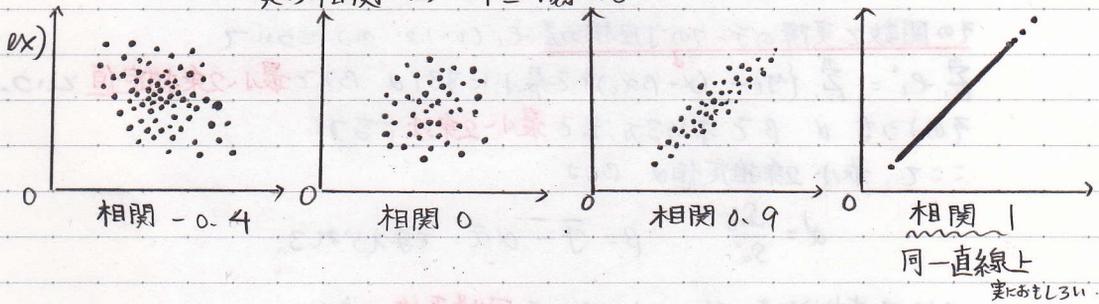
相関の指標で単位のないものを相関係数と言ひ、 r_{xy} で表され、次のように書ける。

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

ここで、正の相関 $\Leftrightarrow 0 < r_{xy} \leq 1$

無相関 $\Leftrightarrow r_{xy} \approx 0$

負の相関 $\Leftrightarrow -1 \leq r_{xy} < 0$



回帰分析

2次元データにおいて x と y の役割が非対称 (x が y を決定する1つの要素となるが、その逆は成り立たない) とき、散布図や相関係数は用いることはできない。

このとき 2変数 x, y の間に x が y を決定、あるいはそれに近い関係があるのなら、

$$y = f(x)$$

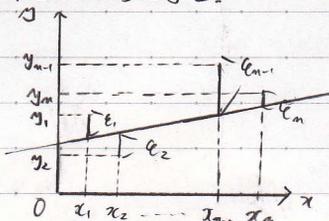
という関数で表せる。

2次元データ $(x_i, y_i) (i=1, 2, \dots, n)$ を用いて $y=f(x)$ がどんな関数なのかを分析する手法を回帰分析という。

と仮定し、 $y_i = \alpha + \beta x_i$ で表せたら最高だが、そんな人ありません☆
なので、その誤差を表す ϵ_i を加えるのだ!!

$$y_i = \alpha + \beta x_i + \epsilon_i \quad (i=1, 2, \dots, n)$$

グラフは下図のようになる。



*直接、 $\alpha, \beta, \epsilon_i$ の値は求めることはできません!!