No.

Date

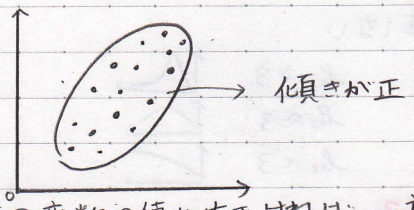
2次元データの要約

2次元データ → (x_1, y_1) (x_2, y_2) ・・・ (x_n, y_n)

2つの種のデータが組み合わされたもの。

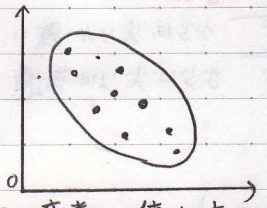
◦散布図

正の相関



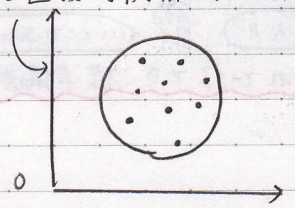
このとき、一方の変数の値が大きければ、一方の変数も大きい。

負の相関



このとき、一方の変数の値が大きければ、一方の変数は小さくなる。

2つのデータが何ら直接的関係を持たないとき、**無相関**であるという。

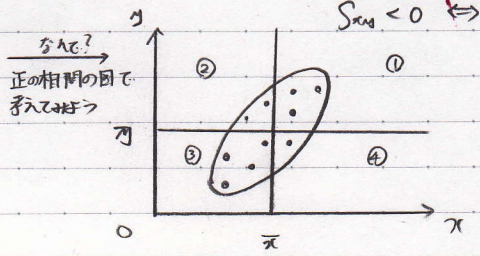


ここで、次式で表される量をxとyの**共分散**と言う

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

ここで、 $S_{xy} > 0 \Leftrightarrow$ 正の相関

$S_{xy} < 0 \Leftrightarrow$ 負の相関 が成立する



①③の領域において、

$(x_i - \overline{x})(y_i - \overline{y})$ は正となり、

②④の領域において、

$(x_i - \overline{x})(y_i - \overline{y})$ は負となるが

この図の場合、$(x_i - \overline{x})(y_i - \overline{y})$ が正となる

データ数の方が多い!!

そうなると $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) > 0$ となり、

これは 正の相関 であると分かり、また、この逆も成立する。