

# Data analysis 5

2つの変量の間の関係を調べたい！

	数 (x)	英 (y)	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x}) \times (y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
A	14	17	2	6	12	4	36
B	10	9	-2	-2	4	4	4
C	13	12	1	1	1	1	1
D	13	10	1	-1	-1	1	1
E	15	12	3	1	3	9	1
F	18	15	6	4	24	36	16
G	6	5	-6	-6	36	36	36
H	10	11	-2	0	0	4	0
I	11	9	-1	-2	2	1	4
J	10	10	-2	-1	2	4	1
計	120	110	0	0	83	100	100

$$\bar{x} = 12, \bar{y} = 11$$

① 偏差の計算.

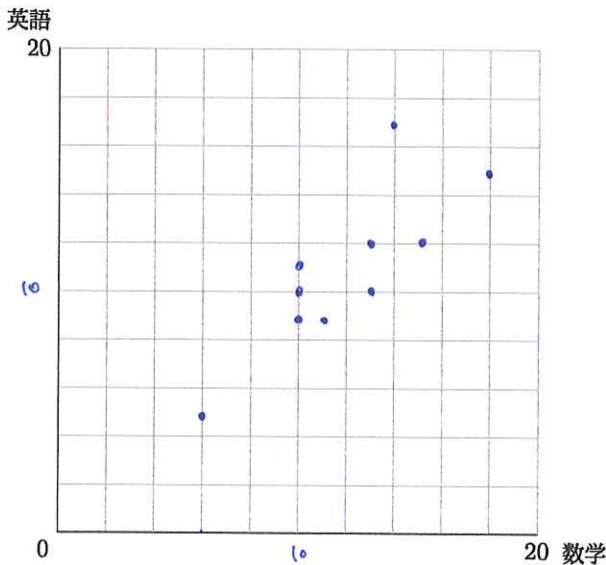
② 分散  $\sigma_x^2 = 10$

$$\sigma_y^2 = 10$$

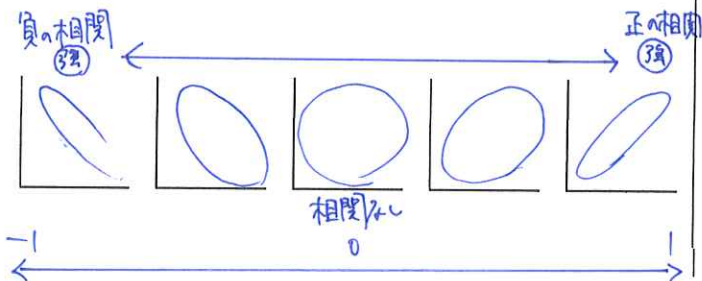
③ 標準偏差

$$\sigma_x = \sqrt{10}, \sigma_y = \sqrt{10}$$

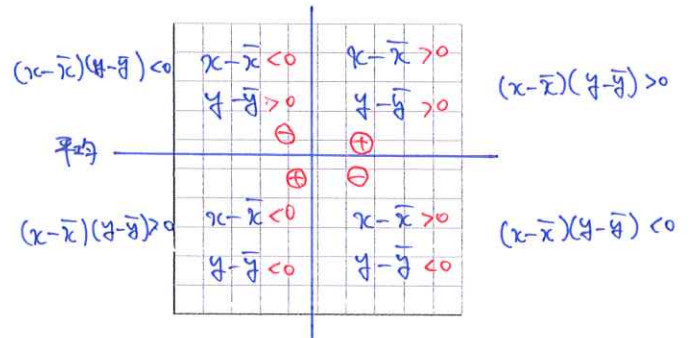
1) 散布図 を描いてみる.



2) 散布図 からみる 相関



2つの変量間の関係を数値で評価したい！



右上、左下傾斜にデータが多い → 正の相関  
右下、左上 " → 負の相関

共分散と相関係数

共分散 →  $x$  と  $y$  の偏差積の平均.

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

相関係数

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$r$  が -1 に近い → 負の相関 が強い.  
 $r$  が 1 に近い → 正の相関

計算してみよう.

3) 共分散

$$\sigma_{xy} = \frac{1}{10} \times 83 = 8.3$$

4) 相関係数

$$r = \frac{8.3}{\sqrt{10} \cdot \sqrt{10}} = 0.83$$

因果関係に??

... 別の原因で、他が起るという関係.

ex.

気温が高いとアイスは売れる。  
等々.

※) 必ずしも「相関」 → 「因果」ではない!!