補充問題 2	共分散と相関係数	CHECK 1	CHECK 2	CHECK 3

次のような 6 組の 2 変数データがある。

$(X, Y) = (6, 7), (14, 3), (8, 6), (18, 1), (2, 9), (12, 4)$

X と Y の標準偏差 S_X, S_Y, および共分散 S_{XY} と相関係数 r_{XY} を求めよ。

ヒント！ 初めからトライ！問題 86 (P130) では，すべてのデータが正の傾きをもつ直線上に並ぶ特殊な場合の問題について解き，そのときの相関係数 $r_{XY}=1$ となることを確認した。今回の問題では，すべての 2 変数データ (X, Y) が負の傾きをもつ直線上に存在する場合について考える。そして，このような場合の相関係数 $r_{XY}=-1$ となることも確認しよう。

解答＆解説

6 組の 2 変数データ

$(X, Y) = (\underbrace{6}_{x_1}, \underbrace{7}_{y_1}), (\underbrace{14}_{x_2}, \underbrace{3}_{y_2}), (\underbrace{8}_{x_3}, \underbrace{6}_{y_3}), (\underbrace{18}_{x_4}, \underbrace{1}_{y_4}), (\underbrace{2}_{x_5}, \underbrace{9}_{y_5}), (\underbrace{12}_{x_6}, \underbrace{4}_{y_6})$ について，

$$\begin{cases} X = x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6 = 6,\ 14,\ 8,\ 18,\ 2,\ 12 \\ Y = y_1,\ y_2,\ y_3,\ y_4,\ y_5,\ y_6 = 7,\ 3,\ 6,\ 1,\ 9,\ 4 \end{cases}$$ とおく。

X と Y の平均値をそれぞれ m_X, m_Y とおき，また X と Y の標準偏差を S_X, S_Y とおき，さらに X と Y の共分散を S_{XY} とおいて，これらを次の表を利用して求める。

平均 $m_X=10$ を引く　　平均 $m_Y=5$ を引く

表

データ No	データ x_k	偏差 x_k-m_X	偏差平方 $(x_k-m_X)^2$	データ y_k	偏差 y_k-m_Y	偏差平方 $(y_k-m_Y)^2$	$(x_k-m_X)(y_k-m_Y)$
1	6	-4	16	7	2	4	$-8\ (=-4\times2)$
2	14	4	16	3	-2	4	$-8\ (=4\times(-2))$
3	8	-2	4	6	1	1	$-2\ (=-2\times1)$
4	18	8	64	1	-4	16	$-32\ (=8\times(-4))$
5	2	-8	64	9	4	16	$-32\ (=-8\times4)$
6	12	2	4	4	-1	1	$-2\ (=2\times(-1))$
合計	60	0	168	30	0	42	-84

前記の表より，

・$\underbrace{m_X}_{X\text{の平均}} = \frac{1}{6}\sum_{k=1}^{6} x_k = \frac{1}{6}\times 60 = 10$ ， $\underbrace{m_Y}_{Y\text{の平均}} = \frac{1}{6}\sum_{k=1}^{6} y_k = \frac{1}{6}\times 30 = 5$

・$\underbrace{S_X{}^2}_{X\text{の分散}} = \frac{1}{6}\sum_{k=1}^{6}(x_k - m_X)^2 = \frac{1}{6}\times 168 = 28$ より，$\underbrace{S_X}_{X\text{の標準偏差}} = \sqrt{28} = 2\sqrt{7}$ ………(答)

$\underbrace{S_Y{}^2}_{Y\text{の分散}} = \frac{1}{6}\sum_{k=1}^{6}(y_k - m_Y)^2 = \frac{1}{6}\times 42 = 7$ より，$\underbrace{S_Y}_{Y\text{の標準偏差}} = \sqrt{7}$ ……………………(答)

・$\underbrace{S_{XY}}_{X\text{と}Y\text{の共分散}} = \frac{1}{6}\sum_{k=1}^{6}(x_k - m_X)(y_k - m_Y) = \frac{1}{6}\times(-84) = -14$ ……………………(答)

以上より，X と Y の相関係数 r_{XY} は，

$r_{XY} = \frac{S_{XY}}{S_X\cdot S_Y} = \frac{-14}{2\sqrt{7}\times\sqrt{7}} = -\frac{14}{14} = -1$ となる。………………………(答)

参考

右図に示すように，今回の 6 組の 2 変数データはすべて，$Y = -\frac{1}{2}X + 10$ という負の傾きをもった直線上に存在している。この場合，相関係数 $r_{XY} = -1$ となるんだね。一般に，相関係数 r_{XY} は，$-1 \leqq r_{XY} \leqq 1$ の範囲の値を取るんだけれど，$r_{XY} = 1$ となるのはすべての 2 変数データが正の傾きの直線上にあるときであり，$r_{XY} = -1$ となるのはすべての 2 変数データが負の傾きの直線上にあるときなんだね。これを，初めからトライ！問題 86 (P130) と，今回の問題で確認したんだね。この結果は覚えておこう！