

We consider the following data points: (2, 19), (9, 6), (7, 15), (5, 12).

a) Calculate the covariance matrix of this set of data.

以下 (1) 和 (2) 均对。在 Excel 等工具中，一般采用 (1)。

(1) 根据样本方差及样本协方差计算得到的协方差矩阵:

We denote the first attribute as  $x$ , and the second attribute as  $y$ .

$$\bar{x} = \frac{2+9+7+5}{4} = 5.75, \quad \bar{y} = \frac{19+6+15+12}{4} = 13$$

$$\sigma_x^2 = \frac{1}{3}[(2-5.75)^2 + (9-5.75)^2 + (7-5.75)^2 + (5-5.75)^2] = 8.917 \text{ (样本方差)}$$

$$\sigma_y^2 = \frac{1}{3}[(19-13)^2 + (6-13)^2 + (15-13)^2 + (12-13)^2] = 30 \text{ (样本方差)}$$

$$\text{cov}(x, y) = \frac{1}{3}[(2-5.75)(19-13) + (9-5.75)(6-13) + (7-5.75)(15-13) + (5-5.75)(12-13)] = -14$$

The covariance matrix is given by  $\begin{bmatrix} 8.917 & -14 \\ -14 & 30 \end{bmatrix}$

上述协方差矩阵，第一行第一列为  $\text{cov}(x, x)$ ；第一行第二列为  $\text{cov}(x, y)$ ；第二行第一列为  $\text{cov}(y, x)$ ；第二行第二列为  $\text{cov}(y, y)$ 。

(2) 根据总体方差及总体协方差计算得到的协方差矩阵:

$$\sigma_x^2 = \frac{1}{4}[(2-5.75)^2 + \dots]$$

$$\sigma_y^2 = \frac{1}{4}[(19-13)^2 + \dots]$$

$$\text{cov}(x, y) = \frac{1}{4}[(2-5.75)(19-13) + \dots]$$

b) Calculate the correlation coefficient between the two attributes.

(1) 根据样本方差及样本协方差计算得到的相关系数：

The correlation coefficient between the two attributes is

$$r_{xy} = \frac{-14}{\sqrt{8.917}\sqrt{30}} = -0.86$$

(2) 根据总体方差及总体协方差计算相关系数，参照课件公式。