

【1】 2つの変量 x と y を持つデータについて、その共分散はどのように求めますか。

【解答】 x と y の共分散 s_{xy} は、次のようにして求めます。

$$s_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

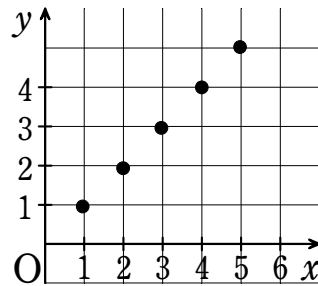
【2】 2つの変量 x と y を持つデータについて、その相関係数はどのように求めますか。

【解答】 x と y の相関係数 r は、次のようにして求めます。 $r = \frac{s_{xy}}{s_x \cdot s_y} \left(= \frac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})} \right)$

【3】 下の表の2つの変量 x と y について、それぞれの標準偏差、共分散、相関係数を求めなさい。

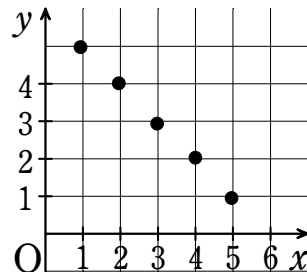
①

x	1	2	3	4	5
y	1	2	3	4	5



②

x	1	2	3	4	5
y	5	4	3	2	1



【解答】 ① x も y も 平均値は3であり、2つの標準偏差は同じといえる。このとき x の標準偏差と y の標準偏差は

$$s_x = s_y = \sqrt{\frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5}} = \sqrt{2} \quad \text{となり、}$$

$$\text{また、共分散は } s_{xy} = \frac{(1-3)(5-3) + (2-3)(4-3) + (3-3)(3-3) + (4-3)(2-3) + (5-3)(1-3)}{5} = 2 \quad \text{となる。}$$

更に相関係数 r は $r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{2}{\sqrt{2} \times \sqrt{2}} = 1$ となる。相関係数が1の時は、図のように x と y の間には完全な正の相関関係がある。

② ①と同様、平均値は3、それぞれの標準偏差も $\sqrt{2}$ となる。しかし、共分散については、

$$s_{xy} = \frac{(1-3)(5-3) + (2-3)(4-3) + (3-3)(3-3) + (4-3)(2-3) + (5-3)(1-3)}{5} = -2 \quad \text{となり、}$$

相関係数 r は、 $r = \frac{-2}{\sqrt{2} \times \sqrt{2}} = -1$ となる。相関係数が -1 の時には、図のように x と y の間には完全な負の相関関係がある。

【4】下の表の2つの変量 x と y について、それぞれの平均値、分散、標準偏差、共分散、相関係数を求めなさい。

x	31	37	30	28	33	27	34	35	29	36
y	22	34	30	20	36	18	38	26	24	32

〔解答〕 まず x について 平均値は $\bar{x} = \frac{31+37+30+28+33+27+34+35+29+36}{10} = 32$

分散は $s_x^2 = \frac{1+25+4+16+1+25+4+9+9+16}{10} = 11$ 標準偏差は $s_x = \sqrt{11}$

次は y について 平均値は $\bar{y} = \frac{22+34+30+20+36+18+38+26+24+32}{10} = 28$

分散は $s_y^2 = \frac{36+36+4+64+64+100+100+4+16+16}{10} = 44$ 標準偏差は $s_y = \sqrt{44} = 2\sqrt{11}$

x と y の共分散については分母が10、分子は

$(-1) \times (-6) + 5 \times 6 + (-2) \times 2 + (-4) \times (-8) + 1 \times 8 + (-5) \times (-10) + 2 \times 10 + 3 \times (-2) + (-3) \times (-4) + 4 \times 4$
 $= 164$ よって 共分散は $s_{xy} = \frac{164}{10} = 16.4$ となる。

このとき 相関係数 $r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{16.4}{\sqrt{11} \times 2\sqrt{11}} = \frac{16.4}{22} \doteq 0.75$

※共分散の求め方は $s_{xy} = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{n}$ で求められます。

但し、 $\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$ は、 $(x_k - \bar{x})(y_k - \bar{y})$ ($k=1, 2, 3, \dots, n$) の総和を表します。