

数学 I 第 5 章 データの分析 No.6

学習のねらい

相関係数について理解しよう！

変量の変換を行った場合の相関係数について理解しよう！

1. 相関係数

共分散で、「相関関係の正 or 負」が分かる。しかし問題がある。まず強弱が分からない。また、比較ができない。というのも、No.5 の例を

例) 以下はお友達の勉強時間とテストの点数をまとめた表である。

お友達	A	B	C	D	E
勉強時間(x時間)	1	2	3	4	5
点数(y点)	40	50	65	70	75

10 点満点にしたら、共分散の値は 100 分の 1 になってしまう。本質は変わらないのに・・・。

そこで、個数や単位に関わらず、上手に-1から1までの範囲を取るようなもの考えた。それが相関係数である。相関係数は r で表す。

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\frac{1}{n} \{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})\}}{\sqrt{\frac{1}{n} \{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}} \sqrt{\frac{1}{n} \{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2\}}}$$

ただし $-1 \leq r \leq 1$ である。

r が1(or-1)に近いとき、強い正(or 負)の相関があるということ。

相関係数を求めるときは、表を書くが良い。先の例で書いてみよう。

生徒	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
A	1	40	-2	-20	4	400	40
B	2	50	-1	-10	1	100	10
C	3	65	0	5	0	25	0
D	4	70	1	10	1	100	10
E	5	75	2	15	4	225	30
計	15	300	0	0	10	850	90

よって、分散 $s_x^2 = \frac{10}{5} = 2$ 、 $s_y^2 = \frac{850}{5} = 170$ となるので、標準偏差 $s_x =$

$\sqrt{2}$ 、 $s_y = \sqrt{170}$ となる。また、共分散 $s_{xy} = \frac{90}{5} = 18$ である。

よって、 $r = \frac{18}{\sqrt{2} \times \sqrt{170}} = 0.976 \dots$ となる。

ゆえに、強い正の相関があると言える。

2. 変量の変換を行った場合の相関係数

$u = ax + b$ 、 $v = cy + d$ で変換するとする。No.4 の公式を使って、 $s_u = |a|s_x$ 、 $s_v = |c|s_y$ 。また、共分散は $s_{uv} = acs_{xy}$ となるので、

$r_{uv} = \frac{ac}{|ac|} \cdot r_{xy}$ となる。 $ac > 0$ のとき、 $r_{uv} = r_{xy}$ 、 $ac < 0$ のとき、 $r_{uv} = -r_{xy}$ 。つまり、変換をしても場合によって符号しか変わらない。

◇問題

1. 次の表は学生 5 名の身長 x (cm)と体重 y (kg)を測定した結果である。
 x と y の相関係数 r を求めよ。

	A	B	C	D	E
身長 x (cm)	181	167	173	169	165
体重 y (kg)	75	59	63	67	61