

# 数学 I 第 5 章 データの分析 No.5

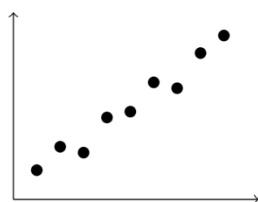
## 学習のねらい

散布図、データの相関関係について理解しよう！  
共分散について理解して公式を覚えよう！

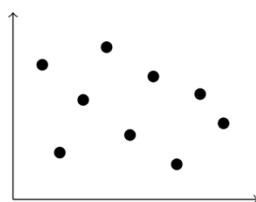
### 1. 散布図、データの相関関係

下の図は、何かしらのデータを座標平面上に、表したものである。(例えば、数学の点数： $x$ 点、英語の点数： $y$ 点など。)このような図を**散布図**という。ここでは形式的に全て横軸を $x$ 軸、縦軸を $y$ 軸と呼ぶことにする。

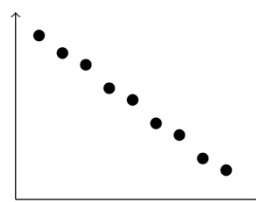
例えば、左の図は、 $x$ が増加するにつれて $y$ も増加している。このような関係を**正の相関関係**があるという。一方、右の図は、 $x$ が増加するにつれて $y$ は減少している。このような関係を**負の相関関係**があるという。真ん中の図のようにどちらの傾向も見られない場合、**相関関係がない**という。



正の相関



どちらとも言えない



負の相関

### 2. 共分散

しかし、散布図を見たときに相関関係の判断に迷った場合……。人によって意見が割れてしまうかもしれない。しかも、毎回散布図を見るのがダルい。つまり「相関関係の正 or 負」が分かれば良いと思うわけだ。そこで用意したのが、**共分散**だ。変数 $x$ 、 $y$ の共分散を $s_{xy}$ という記号で表す。

$$s_{xy} = \frac{1}{n} \{ (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \}$$

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

このデータの共分散を求める。

まず、 $\bar{x} = 3$ 、 $\bar{y} = 60$

生徒	$x$	$y$	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
A	1	40	-2	-20	40
B	2	50	-1	-10	10
C	3	65	0	5	0
D	4	70	1	10	10
E	5	75	2	15	30

$(x - \bar{x})(y - \bar{y})$ の平均は $\frac{1}{5}(40 + 10 + 0 + 10 + 30) = 18$ 。よって、共分散が正なので、勉強時間と点数には正の相関があると考えられる。

## ◇問題

1. 次の表は、ある地域における月ごとのアイスクリーム売り上げと熱中症患者数を表したものである。

月	アイス売り上げ	熱中症患者数
6月	少	少
7月	中	中
8月	多	多

この表を見ると、アイスの売り上げが多い月ほど、熱中症患者数も多いことがわかる。

- (1) この2つのデータには、正の相関・相関なし・負の相関のうちどれが見られるか
- (2) 「アイスを食べると熱中症になる」と結論づけて良いか。理由も説明せよ。