

数学牧場

分散・標準偏差・相関係数

下のような10個のデータ x (平均は $\bar{x} = 6$) があるとき、各データから平均を引いた $x - \bar{x}$ を**偏差**という。偏差はマイナスもあれば、プラスもある。

正負の符号を無視して、単純に「どれくらい平均から離れているか」が知りたいときはマイナスを消すために偏差を2乗した $(x - \bar{x})^2$ を考える。このデータの平均を**分散**という。分散は計算のときに2乗しているため、「平均からどれくらい離れているか」を表すには相応しくないが、 $\sqrt{\quad}$ を取るとそれに近い値が出る。これを**標準偏差**という。

元のデータ
$[x]$ 1, 2, 3, 3, 5, 8, 8, 9, 10, 11
平均は $\bar{x} = 6$
偏差
$[x - \bar{x}]$ -5, -4, -3, -3, -1, 2, 2, 3, 4, 5
偏差の2乗
$[(x - \bar{x})^2]$ 25, 16, 9, 9, 1, 4, 4, 9, 16, 25
平均は 11.8 ... これを分散という。

分散と標準偏差

【分散の求め方】 各データから、平均を引く。→ それぞれ2乗して、また平均を取る。

【標準偏差の求め方】 分散の $\sqrt{\quad}$ を取る。

※標準偏差は「平均からどれだけ離れているか、の平均」を意味する。

問題1 次のデータの分散と標準偏差を求めよ。

- (1) 7点, 7点, 4点, 7点, 8点, 9点, 5点, 9点, 6点, 8点
- (2) 173cm, 176cm, 180cm, 182cm, 184cm

文章問題などの場合、次の公式が便利である。

分散と標準偏差の計算公式

$$\text{分散} = (x^2 \text{の平均}) - (x \text{の平均})^2, \quad \text{標準偏差} = \sqrt{(x^2 \text{の平均}) - (x \text{の平均})^2}$$

問題2 9個の物体の重さを量ろうと思ったが、作業の途中で1個紛失してしまった。残る8個の物体の重さの平均値は200g、標準偏差は15gであった。後日、紛失していた1個が見つかったので、その重さを量ると245gであった。次の問いに答えよ。

- (1) 9個の物体の重さの平均値を求めよ。
- (2) 9個の物体の重さの標準偏差を求めよ。

2種類のデータ x, y がある。それぞれ、平均は $\bar{x} = 4, \bar{y} = 8$ であり、偏差は次のようになる。

$$[x] 3, 4, 5, 4, 4 \quad \rightarrow \quad [x - \bar{x}] -1, 0, 1, 0, 0$$

$$[y] 7, 9, 10, 8, 6 \quad \rightarrow \quad [y - \bar{y}] -1, 1, 2, 0, -2$$

これを利用して、左ページの方法で分散を求めると、 x は0.4、 y は2となる。

ここで、 x と y の偏差どうしをかけたものを考えると、次のようになる。

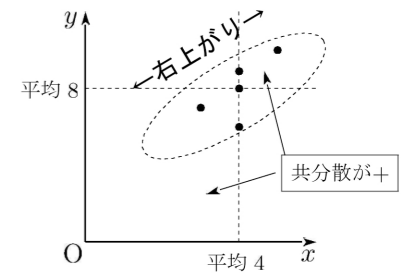
$$[(x - \bar{x})(y - \bar{y})] 1, 0, 2, 0, 0$$

このデータの平均は0.6であるが、これを x と y の**共分散**という。

「 x と y の両方が平均より上か、両方が平均より下」となっているデータが多いとき、共分散は正になる。視覚的に考えると、 x, y を表す点を方眼用紙上にとったとき、平均で4分割した場所の「右上」と「左下」に点がたくさんある状態である。共分散が正ならば、全体として点が「右上がり」分布していることになる。

データが右上がりになっている場合、「 x が大きいと、 y も大きい」傾向があることになる。これを**正の相関がある**という。逆に、データが右下がりになっている場合、「 x が大きいと、 y は小さい」傾向があることになる。これを**負の相関がある**という。

共分散が正なら「正の相関」、共分散が負なら「負の相関」があるといえるが、共分散はデータの種類によって大きさが変わってしまうので、どんな場合でも1よりも小さな値になるように調整したものを、**相関係数**という。



相関係数

$$\text{相関係数 } r = \frac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})} \quad (-1 \leq r \leq 1)$$

1に近いほど**正の相関**があり、-1に近いほど**負の相関**がある。0に近いほど、2つのデータ x, y の間に相関がない。(関係がない)

問題3 ある高校の生徒15名に、ボール投げと懸垂のテストを行った。

ボール投げ (m)	28	25	29	26	28	31	24	27	30	28	27	26	30	32	29
懸垂 (回)	4	7	2	4	7	13	6	8	9	6	2	5	5	5	7

- (1) ボール投げと懸垂の平均をそれぞれ求めよ。
- (2) ボール投げと懸垂の分散をそれぞれ求めよ。
- (3) ボール投げと懸垂の共分散を求めよ。
- (4) ボール投げと懸垂の相関係数を求めよ。また、ボール投げと懸垂の間には相関関係があるといえるか。