数学Ⅰ　第５章　データの分析　No.3

学習のねらい

分散と標準偏差を理解して公式を覚えよう！

1. 分散

例えばＡ組とＢ組、どちらのクラスの点数の散らばりが大きい？

出席番号	1	2	3	4	5
Ａ組	48	49	50	51	52
Ｂ組	20	40	50	60	80

いや、そりゃＢ組でしょ！！！！となると思う。しかしどちらも平均は 50 点である。つまり平均だけだと、散らばり具合が分からない。これだとデータをうまく活用できない。

そこで、平均からの離れ具合(つまり、「(データ)−(平均)」の足し算)を見てみようという発想になったわけだ。このような$x-\overline{x}$のことを**偏差**という。

しかし、これを計算すると、それぞれのデータから 50 を引いて、

Ａ組	−2	−1	0	1	2

となる。これを足すと$-2-1+0+1+2=0$となる。あちゃぁ・・・。ということでボツ！

そこで、全部 2 乗してしまえ！！という発想になった。

Ａ組	4	1	0	1	4

これを足すと、$4+1+0+1+4=10$となる。しかしこのままだとデータの数が多ければ多いほど、値が大きくなってしまう。そこで、「個数で割ろう！！」ということにしたのだ。

よって、Ａ組の分散は$10\div5=2$で2となる。ちなみに同様にＢ組の分散を求めると、400となる。確かにＢ組の方が散らばっているよね、と分かるわけだ。偏差の 2 乗の平均値の値を**分散**といい、s^2で表す。

$$s^2=\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}$$

2. 標準偏差

しかし、分散にも問題がある。「Ｂ組の分散は 400 である！！！」と言ったけど、単位はなんだろう。点数を 2 乗しているから、イメージで言うと、(点数)^{2}となるわけだ。

「えーーと、その・・・。(点数)^{2}ってなんすか？？」

となるよね。その通り。そう思います。そこで、「分散の値に$\sqrt{\ }$を付けちゃえば良いじゃん！！」ってなったわけだ。この値のことを**標準偏差**という。

先の例なら、Ａ組の標準偏差は$\sqrt{2}=1.41\cdots$点、Ｂ組の標準偏差は$\sqrt{400}=20$点となる。実際に表を見てみるとＡ組、Ｂ組の散らばり具合は標準偏差の点数くらいになっている。(平均ジャストの点数の人(3 番の人)がいるので、ちょっと散らばり具合がギュッとなっている感じ。)

公式としてまとめておこう。標準偏差はsで表す。

$$s=\sqrt{\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\}}$$

◊問題

1.　次のデータは、ある商品 A、B の 5 日間の売り上げ個数である。

A：5、7、4、3、6　B：4、6、8、3、9　（単位は個）

A、B の変量をそれぞれx、yとするとき、次の問いに答えよ。

（1）x、yのデータの平均値、分散、標準偏差を求めよ。ただし標準偏差については小数第 2 位を四捨五入せよ。

（2）x、yのデータについて、標準偏差によってどちらの方が散らばり度合いが大きいか判定せよ。