統計学 補助資料

– 算術平均、分散、相関係数についての補足 –

2019 年 5 月 6 日

1 平均の計算

- $x_1, x_2, \cdots, x_n$ のそれぞれに、定数 a を加えたものの算術平均は、 $\bar{x}+a$ (算術平均に a を加える)
- $x_1, x_2, \cdots, x_n$ のそれぞれから、定数 a を引いたものの算術平均は、 $\bar{x}-a$ (算術平均から a を引く)
- $x_1, x_2, \cdots, x_n$ のそれぞれを、定数 a 倍したものの算術平均は、 $a\bar{x}$ (算術平均を a 倍する)
- $x_1, x_2, \cdots, x_n$ のそれぞれを、定数 a で割ったものの算術平均は、 $\frac{\bar{x}}{a}$ (算術平均を a で割る)

簡単な例 $x_1=3, x_2=2, x_3=4$ とすると、$\bar{x}=\dfrac{3+2+4}{3}=\dfrac{9}{3}=3$ である。$a=2$ として、上のそれぞれを考えよう。

- x_1, x_2, x_3 のそれぞれに、2 を加えたものは、$x_1=5, x_2=4, x_3=6$ である。この算術平均は $\dfrac{5+4+6}{3}=5$
これは、$5=3+2$ で、 $\bar{x}+a$
- x_1, x_2, x_3 のそれぞれから、2 を引いたものは、$x_1=1, x_2=0, x_3=2$ である。この算術平均は $\dfrac{1+0+2}{3}=1$
これは、$1=3-2$ で、 $\bar{x}-a$
- x_1, x_2, x_3 のそれぞれを、2 倍したものは、$x_1=6, x_2=4, x_3=8$ である。この算術平均は $\dfrac{6+4+8}{3}=6$
これは、$6=3\times 2$ で、 $a\bar{x}$
- x_1, x_2, x_3 のそれぞれを、2 で割ったものは、$x_1=\frac{3}{2}, x_2=1, x_3=2$ である。この算術平均は $\dfrac{\frac{3}{2}+1+2}{3}=\dfrac{3}{2}$
これは、$\frac{3}{2}=3\div 2$ で、 $\frac{\bar{x}}{a}$

2 分散の計算

- $x_1, x_2, \cdots, x_n$ のそれぞれに、定数 a を加えたものの分散は s^2 (分散は変わらない)
- $x_1, x_2, \cdots, x_n$ のそれぞれに、定数 a を加えたものの分散は s^2 (分散は変わらない)
- $x_1, x_2, \cdots, x_n$ のそれぞれを、定数 a 倍したものの分散は a^2s^2 (分散の a^2 倍)
- $x_1, x_2, \cdots, x_n$ のそれぞれを、定数 a で割ったものの分散は $\frac{s^2}{a^2}$ (分散を a^2 で割る)

簡単な例 $x_1=3, x_2=2, x_3=4$ とすると、$s^2=\dfrac{(3-3)^2+(2-3)^2+(4-3)^2}{3}=\dfrac{0+1+1}{3}=\dfrac{2}{3}$ である。$a=2$ として、上のそれぞれを考えよう。

- x_1, x_2, x_3 のそれぞれに、2 を加えたものは、$x_1=5, x_2=4, x_3=6$ である。この算術平均は 5 であり、分散は $s^2=\dfrac{(5-5)^2+(4-5)^2+(6-5)^2}{3}=\dfrac{0+1+1}{3}=\dfrac{2}{3}$ これは、元の分散と同じである。
- x_1, x_2, x_3 のそれぞれから、2 を引いたものは、$x_1=1, x_2=0, x_3=2$ である。この算術平均は 1 であり、分散は $s^2=\dfrac{(1-1)^2+(0-1)^2+(2-1)^2}{3}=\dfrac{0+1+1}{3}=\dfrac{2}{3}$ これは、元の分散と同じである。
- x_1, x_2, x_3 のそれぞれを、2 倍したものは、$x_1=6, x_2=4, x_3=8$ である。この算術平均は 6 であり、分散は $s^2=\dfrac{(6-6)^2+(4-6)^2+(8-6)^2}{3}=\dfrac{0+4+4}{3}=\dfrac{8}{3}$ これは、$\frac{8}{3}=\frac{2}{3}\times 2^2$ で、a^2s^2
- x_1, x_2, x_3 のそれぞれを、2 で割ったものは、$x_1=\frac{3}{2}, x_2=1, x_3=2$ である。この算術平均は $\frac{3}{2}$ であり、分散は $s^2=\dfrac{(\frac{3}{2}-\frac{3}{2})^2+(1-\frac{3}{2})^2+(2-\frac{3}{2})^2}{3}=\dfrac{0+\frac{1}{4}+\frac{1}{4}}{3}=\dfrac{1}{6}$ これは、$\frac{1}{6}=\frac{2}{3}\div 2^2$ で、$\frac{s^2}{a^2}$

3　相関係数の計算

$(x_1, x_2, \cdots, x_n)$ と $(y_1, y_2, \cdots, y_n)$ の相関係数は、x と y のデータを入れ替えたり、個々のデータをすべて a 倍したりしても、変わらない。

以下では、第 12 回練習問題で用いた数値例について、これらのことを示してみる。なお、元のデータの相関係数は、

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{9900}{\sqrt{11000 \times 11000}} = \frac{9900}{11000} = 0.9$$

である。

x と y を入れ替えた例

番号	英語 (x)	数学 (y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	50	60	0	10	0	100	0
2	20	0	-30	-50	900	2500	1500
3	70	80	20	30	400	900	600
4	0	10	-50	-40	2500	1600	2000
5	40	40	-10	-10	100	100	100
6	30	20	-20	-30	400	900	600
7	60	50	10	0	100	0	0
8	90	70	40	20	1600	400	800
9	100	90	50	40	2500	1600	2000
10	10	30	-40	-20	1600	400	800
11	80	100	30	50	900	2500	1500
平均	50	50		計	11000	11000	9900

この相関係数は、

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{9900}{\sqrt{11000 \times 11000}} = \frac{9900}{11000} = 0.9$$

である。

x と y の個々のデータをそれぞれ **100** 倍した例

番号	英語 (x)	数学 (y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	6000	5000	1000	0	1000000	0	0
2	0	2000	-5000	-3000	25000000	9000000	15000000
3	8000	7000	3000	2000	9000000	4000000	6000000
4	1000	0	-4000	-5000	16000000	25000000	20000000
5	4000	4000	-1000	-1000	1000000	1000000	1000000
6	2000	3000	-3000	-2000	9000000	4000000	6000000
7	5000	6000	0	1000	0	1000000	0
8	7000	9000	2000	4000	4000000	16000000	8000000
9	9000	10000	4000	5000	16000000	25000000	20000000
10	3000	1000	-2000	-4000	4000000	16000000	8000000
11	10000	8000	5000	3000	25000000	9000000	15000000
平均	5000	5000		計	110000000	110000000	99000000

この相関係数は、

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{99000000}{\sqrt{110000000 \times 110000000}} = \frac{99000000}{110000000} = 0.9$$

である。

また、片側だけ 100 倍したとしても、相関係数は変わらない。そのほか、一方または両方の個々のデータにすべて定数 a を加えたり、引いたりしても相関係数は変わらない。