

雑感

標準偏差が整数のデータ (続)

■ 雑感 99 で「標準偏差が整数のデータ」を載せたが、データ数が 10 のもののセットだから、組み合わせで用いる場合、10 の倍数のデータに限られる。

半端な数のデータにはどう対応したらいいか。その都度、(EXCEL か何かで確認しながら調整して) 作ってしまえばいいのだが、他に方法はないか。

■ 偶数個にするには妙手がある。

雑感 93 「データの分析のオリジナル問題」に載せた問題の(2)の結果が使える。

標本数 $n (\geq 2)$ のデータ $X = \{x_1, x_2, x_3, \dots, x_n\}$ の平均値が m 、標準偏差が $\sigma (> 0)$ であるとする。

(2) X に 2 つの標本 y_1, y_2 ($y_1 \geq y_2$) を付け加えたデータで、平均値も標準偏差も変わらないとき、 y_1, y_2 を求めよ。

結果は、 $y_1 = m + \sigma, y_2 = m - \sigma$ であり、雑感 93 にも書いたように、「データ X に $m + \sigma$ の値を持つ標本を p 個と $m - \sigma$ の値を持つ標本を p 個付け加えても、平均値も標準偏差も変わらない」のである。

したがって、雑感 99 に載せたデータ数 10、平均値 0 のデータに対して、そのデータの標準偏差 σ に対して、値 σ と値 $-\sigma$ のデータを同数付け加えればいいのだ。

例えば、データ数 12、標準偏差 3、平均値 0 のデータが必要ならば、データ $\{4, 3, 3, 2, 1, 0, -1, -3, -4, -5\}$ (分散 9) に、3、 -3 のデータを付け加えて

$\{4, 3, 3, 3, 2, 1, 0, -1, -3, -3, -4, -5\}$

というデータを作ればいいのである。

■ しかし、奇数個にするにはこの手は使えない。したがって、元になる奇数個のデータが必要になる。

EXCEL で調整しながら、実験的に標本数 9 のデータを作ったものが、下の表である。

分散の値が偶数に限られているのは、致し方ないことである。

実際、平均値 0 から $\sum_{k=1}^9 x_k = 0$ で、

$$9\sigma^2 = \sum_{k=1}^9 x_k^2 - 0^2 = \left(\sum_{k=1}^9 x_k\right)^2 - 2\sum_{i \neq j} x_i x_j = -2\sum_{i \neq j} x_i x_j$$

であり、 $x_k (1 \leq k \leq 9)$ 、 σ^2 が整数であるならば、 σ^2 は偶数に限られるのである。

これを元にして、B, H, A \cup B \cup E などのデータに偶数個のデータを先程と同じように付け加えていけば、整数の標準偏差を持つ奇数個のデータを作ることができる。

No	A	B	C	D	E	F	G	H	I
	2	3	4	4	5	5	5	5	5
	2	2	2	3	4	4	5	4	5
	1	1	2	2	2	3	4	4	4
	0	1	1	1	1	2	1	4	4
	0	0	0	0	0	0	0	-1	-1
	0	0	0	0	-1	-2	-3	-2	-2
	-1	-1	-2	-1	-3	-3	-3	-4	-5
	-2	-2	-3	-4	-3	-4	-4	-5	-5
	-2	-4	-4	-5	-5	-5	-5	-5	-5
sum	0	0	0	0	0	0	0	0	0
ave	0	0	0	0	0	0	0	0	0
var	2	4	6	8	10	12	14	16	18

なお、老婆心ながら付言すれば、雑感 99 のデータとこのデータを組み合わせで用いることはできない。それは、データ数が 10 と 9 のように異なるからである。