

記述統計 各種統計量の定義と求め方

- 1) 平均
- 2) 分散
- 3) 標準偏差

- 1) 平均

X を確率変数、 x_1, \dots, x_n を標本とするとき、

$$E(X) = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

を標本平均と呼ぶ。

- 2) 分散

$$V(X) = \frac{\sum_{i=1}^n (x_i - E(X))^2}{n} \quad (2)$$

を標本分散と呼ぶ

(注意) 分散を数値的に計算するときに、(2)式を変形した、

$$V(X) = \frac{\sum_{i=1}^n (x_i - E(X))^2}{n} = E(X^2) - \{E(X)\}^2$$

の公式を単純に用いてはならない。 $E(X^2)$ の計算の時に、数値的な誤差(桁落ち)が起きて、真の値よりも得られた値が小さくなる可能性がある。

例題

以下のデータの分散を(2)式で求めると負になる！

解決方法1 数値表現の精度を上げる(倍精度、4倍精度など)

実際にはこれで解決するケースが多いが、システムティックではない。

解決方法2 仮の平均を計算して、その平均をデータから引いて(2)式で分散を計算する。うまくいくケースが多い。

解決方法2については以下の表を参照すること。

x_i	x_i^2	$(x_i - E(X))^2$	$(x_i - \mu)$	$(x_i - \mu)^2$
1.00080	1.00160	2.9241E-08	0.00080	6.40000E-07
1.00034	1.00067	8.3521E-08	0.00034	1.15600E-07
1.00057	1.00114	3.481E-09	0.00057	3.24900E-07
1.00071	1.00141	6.561E-09	0.00071	5.04100E-07
1.00093	1.00186	9.0601E-08	0.00093	8.64900E-07
1.00066	1.00132	9.61E-10	0.00066	4.35600E-07
1.00045	1.00089	3.2041E-08	0.00045	2.02500E-07
1.00035	1.00069	7.7841E-08	0.00035	1.22500E-07
1.00095	1.00190	1.03041E-07	0.00095	9.02500E-07
1.00053	1.00106	9.801E-09	0.00053	2.80900E-07
$E(x)$	$E(x^2)$	μ (仮平均)	$E(X-\mu)$	$E((X-\mu)^2)$
1.00063	1.001254	1	0.00063	4.3935E-07
$E(x^2) - (E(x))^2$	$E((X-E(X))^2)$		$E((X-\mu)^2) - (E(X-\mu))^2$	
-6.3969E-06	4.3709E-08		4.371E-08	