計算の工夫をして平均と分散を求める

コンピュータを使うと，数値が大きくても瞬時に平均や分散を求められるので，計算方法を工夫しないが，コンピュータを使わずに計算する場合，一工夫することで計算量を大幅に減らすことができる。

その方法とは，すべての数値から一定の値を引いて，数値の絶対値をできるだけ下げ，途中の計算を簡単にして平均や分散を求める方法である。

数値を下げて計算した仮平均に，全体から引いた一定値を後で加えることで、求める平均値が得られる※1。

分散は,分布の広がりだから,本来の状態で求めても,全体的に一定値を引いて求めても，分布の広がりは変わらないので，数値を下げて得られた分散がそのまま答えとなる※2。

次の例題を使って計算量を比較する。

［例題］次のデータを用いて平均(m)と分散(V)の値を求めよ。

	1	2	3	4	5	6	7	8	9	10
X	62	57	77	48	69	58	81	53	64	74

（平均）

＜一般的な方法＞

$$平均=\frac{62+57+77+\cdots+53+64+74}{10}=\frac{643}{10}=64.3$$

＜数値を下げて求める方法＞

数値を見て，だいたい 60 あたりに集中しているから，全体から 60 を引いて平均を求める。

	1	2	3	4	5	6	7	8	9	10
Y	2	−3	17	−12	9	−2	21	−7	4	14

$$仮平均=\frac{2+(-3)+17+\cdots+(-7)+4+14}{10}=\frac{43}{10}=4.3$$

よって，平均は，64.3

（注）・全体から引く数字は，データから推測すればよい。

・一般的な方法に比べて，数値がかなり小さくなり，計算しやすい。

（分散）

＜一般的な方法①＞・・・$\sum(X_i-m)^2$

	1	2	3	4	5	6	7	8	9	10
X	62	57	77	48	69	58	81	53	64	74
$(X-m)^2$	5.29	53.29	161.29	265.69	22.09	39.69	278.89	127.69	0.09	94.09

$$分散=\frac{5.29+53.29+161.29+\cdots+127.69+0.09+94.09}{10}=\frac{1048.1}{10}=104.81$$

（注）・各データから，平均を引いて小数の２乗をするので，手計算ではかなり手間がかかる。

＜一般的な方法②＞・・・ $E(X^2)-(E(X))^2$

	1	2	3	4	5	6	7	8	9	10
X	62	57	77	48	69	58	81	53	64	74
X²	3844	3249	5929	2304	4761	3364	6561	2809	4096	5476

よって，X^2の平均は

$$X^2\text{の平均}=\frac{3844+3249+5929+\cdots+2809+4096+5476}{10}=\frac{42393}{10}=4239.3$$

分散＝4239.3－(64.3)^2＝104.81

（注）・小数の２乗計算がなくなった分，楽になったが，X^2の平均を求めるので，大きな数字になってしまう。

＜数値を下げて求める方法＞・・・ $E(Y^2)-(E(Y))^2$

	1	2	3	4	5	6	7	8	9	10
Y	2	−3	17	−12	9	−2	21	−7	4	14
Y²	4	9	289	144	81	4	441	49	16	196

よって，Y^2の平均は

$$Y^2\text{の平均}=\frac{4+9+289+\cdots+49+16+196}{10}=\frac{1233}{10}=123.3$$

分散＝123.3－(4.3)^2＝104.81

（注）・扱う数値が小さいので，Y^2の平均を求める時も比較的平易である。

※１　確率変数 X_i から，一定値 m を引いた値を Y_i とする。

$$Y_i=X_i-m \qquad \therefore \quad X_i=Y_i+m$$

このとき，X_i の平均 $E(X)$ は

$$\begin{aligned}E(X)&=\frac{1}{N}\sum_{k=1}^{N}X_k\\&=\frac{1}{N}\sum_{k=1}^{N}(Y_k+m)\\&=\frac{1}{N}\sum_{k=1}^{N}Y_k+\frac{1}{N}\sum_{k=1}^{N}m\\&=E(Y)+m\end{aligned}$$

したがって，X_i の平均を求めるとき，一定値 m を引いた値 Y_i の平均を求めておき，最後に一定値 m 加えてもよいことが分かる。このときの $E(Y)$ を仮平均という。

※２　確率変数 X_i から，一定値 m を引いた値 Y_i の分散 $V(Y)$ は。

$$\begin{aligned}V(Y)&=E(Y^2)-\{E(Y)\}^2\\&=\frac{1}{N}\sum_{k=1}^{N}(X_i-m)^2-\{E(X)-m\}^2\\&=\frac{1}{N}\sum_{k=1}^{N}X_i^{\,2}+\frac{2}{N}\sum_{k=1}^{N}mX_i+\frac{1}{N}\sum_{k=1}^{N}m^2-\left\{\{E(X)\}^2+2m\cdot E(X)+m^2\right\}\\&=E(X^2)+2m\cdot E(X)+m^2-\left\{\{E(X)\}^2+2m\cdot E(X)+m^2\right\}\\&=E(X^2)-\{E(X)\}^2=V(X)\end{aligned}$$