

# 1 変数変換による平均値, 不偏分散, 標準偏差

測定値を  $X_i$  とする ( $i = 1, 2, \dots, n$ )。以下のように新しい変数  $A, B, C$  を定義したとき, 各変数の平均値, 不偏分散, 標準偏差は元の変数  $X$  の平均値, 不偏分散, 標準偏差とどのような関係になるか。

- (a)  $A_i = a X_i, a \neq 0$
- (b)  $B_i = X_i + b$
- (c)  $C_i = a X_i + b, a \neq 0$

変数  $X, A, B, C$  の平均値, 不偏分散, 標準偏差をそれぞれ,  $\bar{X}, U_X, SD_X ; \bar{A}, U_A, SD_A ; \bar{B}, SD_B ; \bar{C}, U_C, SD_C$  とする<sup>注1</sup>。

(a)

$$\begin{aligned}\bar{A} &= \frac{\sum_{i=1}^n A_i}{n} = \frac{\sum_{i=1}^n a X_i}{n} = \frac{a \sum_{i=1}^n X_i}{n} = a \bar{X} \\ U_A &= \frac{\sum_{i=1}^n (A_i - \bar{A})^2}{n-1} = \frac{\sum_{i=1}^n (a X_i - a \bar{X})^2}{n-1} = \\ &= \frac{\sum_{i=1}^n a^2 (X_i - \bar{X})^2}{n-1} = \frac{a^2 \sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = a^2 U_X \\ SD_A &= \sqrt{U_A} = \sqrt{a^2 U_X} = a \sqrt{U_X} = a SD_X\end{aligned}$$

(b)

$$\begin{aligned}\bar{B} &= \frac{\sum_{i=1}^n B_i}{n} = \frac{\sum_{i=1}^n (X_i + b)}{n} = \frac{\sum_{i=1}^n X_i + \sum_{i=1}^n b}{n} \\ &= \frac{\sum_{i=1}^n X_i + n b}{n} = \frac{\sum_{i=1}^n X_i}{n} + \frac{n b}{n} = \bar{X} + b \\ U_B &= \frac{\sum_{i=1}^n (B_i - \bar{B})^2}{n-1} = \frac{\sum_{i=1}^n \{(X_i + b) - (\bar{X} + b)\}^2}{n-1} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = U_X \\ SD_B &= \sqrt{U_B} = \sqrt{U_X} = \sqrt{U_X} = SD_X\end{aligned}$$

(c)

$$\begin{aligned}\bar{C} &= \frac{\sum_{i=1}^n C_i}{n} = \frac{\sum_{i=1}^n (a X_i + b)}{n} = \frac{\sum_{i=1}^n a X_i + \sum_{i=1}^n b}{n} \\ &= \frac{a \sum_{i=1}^n X_i + n b}{n} = a \frac{\sum_{i=1}^n X_i}{n} + \frac{n b}{n} = a \bar{X} + b\end{aligned}$$

---

<sup>注1</sup>  $\Sigma$  記号を伴う演算については、3 ページを参照せよ。

$$\begin{aligned}
U_C &= \frac{\sum_{i=1}^n (C_i - \bar{C})^2}{n-1} = \frac{\sum_{i=1}^n \{(aX_i + b) - (a\bar{X} + b)\}^2}{n-1} \\
&= \frac{\sum_{i=1}^n (aX_i - a\bar{X})^2}{n-1} = \frac{\sum_{i=1}^n a^2(X_i - \bar{X})^2}{n-1} \\
&= a^2 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = a^2 U_X \\
SD_C &= \sqrt{U_C} = \sqrt{a^2 U_X} = a \sqrt{U_X} = a SD_X
\end{aligned}$$

確かめてみるのは簡単である。

1.  $X$  として、1, 2, 3 を考える ( $n = 3$ 。データは何個でも、どんな数でもよい)。

$$\begin{aligned}
\star \bar{X} &= (1+2+3)/3 = 2 \\
\star U_X &= \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3-1} = 1
\end{aligned}$$

2. (c) を考えてみる。 $a = 3, b = 5$  とすると、 $C$  は 8, 11, 14 である。

$$\begin{aligned}
\star \bar{C} &= (8+11+14)/3 = 11 = a\bar{X} + b \\
\star U_C &= \frac{(8-11)^2 + (11-11)^2 + (14-11)^2}{3-1} = 9 = a^2 U_X
\end{aligned}$$

## 2 シグマ記号

統計学の教科書には、 $\sum_{i=1}^n X_i$ というような、Σ記号が入った数式がよく出てくる。その意味は「式  $X_i$  の  $i$  を 1 から始めて 1 ずつ増加させ、 $i$  が  $n$  になるまで、 $X_i$  の和をとる」ことである。つまり、

$$\sum_{i=1}^n X_i = X_1 + X_2 + \cdots + X_n$$

のことである。

和を取るものと表す式（上の例では  $X_i$ ）が複雑な形をしていても、本来の意味を表すように書き改めると理解しやすい。

例 1 :  $a$  が定数で、 $X_1, X_2, \dots, X_n$  があるとき、

$$\sum_{i=1}^n aX_i = aX_1 + aX_2 + \cdots + aX_n = a(X_1 + X_2 + \cdots + X_n) = a \sum_{i=1}^n X_i$$

例 2 :  $n$  個のデータ、 $X_1, X_2, \dots, X_n$  があるとき、平均値は以下のように定義される。

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

例 3 :  $n$  個のデータ、 $X_1, X_2, \dots, X_n$  があるとき、分散は以下のように定義される。例 2 で示した平均の定義も利用して、左辺から右辺を導く。

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

答え :

$$\begin{aligned} V &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{1}{n} \left\{ (X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2 \right\} \\ &= \frac{1}{n} \left\{ (X_1^2 - 2X_1\bar{X} + \bar{X}^2) + \cdots + (X_n^2 - 2X_n\bar{X} + \bar{X}^2) \right\} \\ &= \frac{1}{n} \left\{ (X_1^2 + \cdots + X_n^2) - 2\bar{X}(X_1 + \cdots + X_n) + (\bar{X}^2 + \cdots + \bar{X}^2) \right\} \\ &= \frac{1}{n} \left\{ \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \right\} \\ &= \frac{\sum_{i=1}^n X_i^2}{n} - 2\bar{X}^2 + \bar{X}^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 \end{aligned}$$