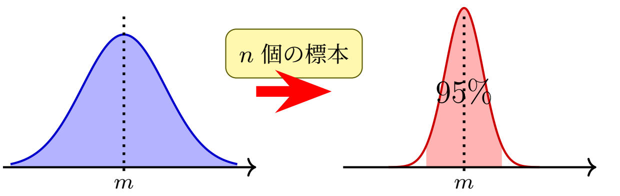
母標準偏差 σ が既知の場合、$\sigma(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$ であるが、実際の調査では σ は未知である。
では、標本から計算した標準偏差で σ を代用できるのか？
この２つはどの程度一致するのだろうか？

母集団 X の分布　　　　**標本平均 $\overline{X}$ の分布**



母平均：$\mathrm{E}(X) = m$
母標準偏差：$\sigma(X) = \sigma$

標本平均の期待値：$\mathrm{E}(\overline{X}) = m$
標本平均の標準偏差：$\sigma(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$

まず、小さな有限母集団で実際に計算してみよう。

例 1　母集団 1,3,5,7 から 3 つの数字を非復元抽出で取り出すとき、次の問いに答えよ。

(1) 母集団の平均 m と標準偏差 σ を求めよ。

答　$m =$　　　　$\sigma =$

(2) 可能なすべての標本について、標本平均 $\bar{x}$ と標本標準偏差 s_0 を求めよ（ただし分母は n）。

$$s_0 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$\{1,3,5\}$ のとき、

$\{1,3,7\}$ のとき、

$\{1,5,7\}$ のとき、

$\{3,5,7\}$ のとき、

(3) 標本標準偏差 s_0 の確率分布（各値とその確率）を求めよ。

(4) 標本標準偏差 s_0 の期待値（平均値）を求めよ。

$E(s_0) =$

(5) (1) と (4) の結果を比較し、気づいたことを述べよ。

なぜそうなるのか？一般的な場合（無限母集団）で理論的に考察しよう。

母集団から n 個の標本 $X_1, X_2, \ldots, X_n$ を取り出したとき、標本分散（分母 n）を

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

とする。ただし、$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ は標本平均である。
この標本分散の期待値を求めると：

$$\mathrm{E}(s_0^2) = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

ここで、

$$\begin{aligned}
&\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \sum_{i=1}^{n}\{(X_i - m) - (\bar{X} - m)\}^2 \\
&= \sum_{i=1}^{n}\left\{(X_i - m)^2 - 2(X_i - m)(\bar{X} - m) + (\bar{X} - m)^2\right\} \\
&= \sum_{i=1}^{n}(X_i - m)^2 - 2(\bar{X} - m)\sum_{i=1}^{n}(X_i - m) + \sum_{i=1}^{n}(\bar{X} - m)^2 \\
&= \sum_{i=1}^{n}(X_i - m)^2 - 2(\bar{X} - m)\cdot n(\bar{X} - m) + n(\bar{X} - m)^2 \\
&= \sum_{i=1}^{n}(X_i - m)^2 - 2n(\bar{X} - m)^2 + n(\bar{X} - m)^2 \\
&= \sum_{i=1}^{n}(X_i - m)^2 - n(\bar{X} - m)^2
\end{aligned}$$

したがって

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

期待値の和の性質を利用すると、

$$
\begin{aligned}
\mathrm{E}(s_0^2) &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2-(\bar{X}-m)^2\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2\right]-\mathrm{E}\left[(\bar{X}-m)^2\right] \\
&= \sigma^2-\mathrm{V}(\bar{X}) \\
&= \sigma^2-\frac{\sigma^2}{n} \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

結論：

$$\boxed{\mathrm{E}(s_0^2)=\frac{n-1}{n}\sigma^2<\sigma^2}$$

分母 n の標本分散の期待値は母分散よりも $\frac{n-1}{n}$ 倍だけ小さくなる。
母分散を正確に推定するには、標本分散を $\frac{n}{n-1}$ 倍する。
これを**不偏分散**といい、不偏分散の平方根を**標本標準偏差**という。母平均の推定の場合はこれを用いる。

標本標準偏差

母標準偏差を推定するための標本標準偏差

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

※ 無限母集団からの復元抽出を想定

※ 分母が n ではなく $(n-1)$ であることに注意

※ s^2 を**不偏分散**という

不偏分散の分母が $n-1$ になる理由の一つは、**偏差の自由度**の考え方による。
n 個のデータから標本平均 $\bar{x}$ を計算すると、偏差 $(x_i-\bar{x})$ には

$$\sum_{i=1}^{n}(x_i-\bar{x})=0$$

という制約条件が生まれる。この制約により、n 個の偏差のうち独立な情報を持つのは $(n-1)$ 個だけとなる。
例： $n=3$ のとき、2 つの偏差が決まれば、残りの 1 つは制約条件から自動的に決まる。
この独立な情報の個数を**自由度（degrees of freedom）**といい、偏差平方和 $\sum(x_i-\bar{x})^2$ の自由度は $(n-1)$ である。
したがって、不偏分散は

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

のように、自由度で割ることで正しい推定量となる。

問 1　標本 $\{2,5,8\}$ が得られた。

(1) 標本平均 $\bar{x}$ を求めよ。

答　$\bar{x}=$

(2) 標本分散（分母 n）$s_0^2=\frac{1}{n}\sum(x_i-\bar{x})^2$ を求めよ。

答　$s_0^2=$

(3) 不偏分散（分母 $n-1$）$s^2=\frac{1}{n-1}\sum(x_i-\bar{x})^2$ を求めよ。

答　$s^2=$

問 2　標本 $\{3,7,x_3\}$ が得られ、標本平均は $\bar{x}=7$ である。

(1) 2 つの偏差 $(x_1-\bar{x}),(x_2-\bar{x})$ を求めよ。

答

(2) $(x_3-\bar{x})$ を求めよ。

答

(3) この結果から、「自由度が 2」の意味を説明せよ。

答

++*+*+*+*+*+*+ **【解答】** *+*+*+*+*+*+*+*+

例 1 母集団 1,3,5,7 から 3 つの数字を非復元抽出で取り出すとき、次の問いに答えよ。

(1) 母集団の平均 m と標準偏差 σ を求めよ。

$m = \frac{1+3+5+7}{4} = 4$

$\sigma^2 = \frac{(1-4)^2+(3-4)^2+(5-4)^2+(7-4)^2}{4} = 5$

$\sigma = \sqrt{5} \approx 2.236$

答 $m = 4, \sigma = \sqrt{5}$

(2) 可能なすべての標本について、標本平均 $\bar{x}$ と標本標準偏差 s_0（分母 n）を求めよ。

$$s_0 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$\{1,3,5\}$ のとき、

$\bar{x} = 3, \quad s_0 = \sqrt{\frac{(1-3)^2+(3-3)^2+(5-3)^2}{3}} = \sqrt{\frac{8}{3}}$

$\{1,3,7\}$ のとき、

$\bar{x} = \frac{11}{3}, \quad s_0 = \sqrt{\frac{(1-\frac{11}{3})^2+(3-\frac{11}{3})^2+(7-\frac{11}{3})^2}{3}} = \sqrt{\frac{56}{9}}$

$\{1,5,7\}$ のとき、

$\bar{x} = \frac{13}{3}, \quad s_0 = \sqrt{\frac{(1-\frac{13}{3})^2+(5-\frac{13}{3})^2+(7-\frac{13}{3})^2}{3}} = \sqrt{\frac{56}{9}}$

$\{3,5,7\}$ のとき、

$\bar{x} = 5, \quad s_0 = \sqrt{\frac{(3-5)^2+(5-5)^2+(7-5)^2}{3}} = \sqrt{\frac{8}{3}}$

(3) 標本標準偏差 s_0 の確率分布（各値とその確率）を求めよ。

s_0	$\sqrt{\frac{8}{3}}$	$\sqrt{\frac{56}{9}}$
$P(s_0)$	$\frac{1}{2}$	$\frac{1}{2}$

(4) 標本標準偏差 s_0 の期待値（平均値）を求めよ。

$E(s_0) = \sqrt{\frac{8}{3}} \times \frac{1}{2} + \sqrt{\frac{56}{9}} \times \frac{1}{2}$

$= \frac{1}{2}\left(\sqrt{\frac{8}{3}} + \sqrt{\frac{56}{9}}\right) \approx 2.076$

(5) (1) と (4) の結果を比較し、気づいたことを述べよ。

母標準偏差 $\sigma = \sqrt{5} \approx 2.236$ に対し、標本標準偏差の期待値 $E(s_0) \approx 2.076$ は小さい。

つまり、分母 n の標本標準偏差は母標準偏差を**過小評価**する傾向がある。

問 1 標本 $\{2,5,8\}$ が得られた。

(1) 標本平均 $\bar{x}$ を求めよ。

$$\bar{x} = \frac{2+5+8}{3} = 5$$

答 $\bar{x} = 5$

(2) 標本分散（分母 n）$s_0^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$ を求めよ。

$$s_0^2 = \frac{(2-5)^2+(5-5)^2+(8-5)^2}{3} = \frac{18}{3} = 6$$

答 $s_0^2 = 6$

(3) 不偏分散（分母 $n-1$）$s^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2$ を求めよ。

$$s^2 = \frac{(2-5)^2+(5-5)^2+(8-5)^2}{3-1} = \frac{18}{2} = 9$$

答 $s^2 = 9$

問 2 標本 $\{3,7,x_3\}$ が得られ、標本平均は $\bar{x} = 7$ である。

(1) 2 つの偏差 $(x_1-\bar{x}), (x_2-\bar{x})$ を求めよ。

$$(x_1-\bar{x}) = 3-7 = -4$$

$$(x_2-\bar{x}) = 7-7 = 0$$

答 $-4, 0$

(2) $(x_3-\bar{x})$ を求めよ。

偏差の和は 0 なので、

$$(x_1-\bar{x}) + (x_2-\bar{x}) + (x_3-\bar{x}) = 0$$

$$-4+0+(x_3-\bar{x}) = 0$$

$$(x_3-\bar{x}) = 4$$

答 4

(3) この結果から、「自由度が 2」の意味を説明せよ。

3 つの偏差のうち、2 つの値が決まれば、残りの 1 つは制約条件 $\sum(x_i-\bar{x}) = 0$ から自動的に決まる。

したがって、独立に決められる偏差の個数は $(3-1) = 2$ 個である。

これを「自由度が 2」という。

答 上記参照