

『データ分析をマスターする 12 のレッスン 〔新版〕』

畑農鋭矢・水落正明 〔著〕

補論

発行所 株式会社有斐閣

2022 年 12 月 10 日 初版第 1 刷発行

ISBN 978-4-641-22205-1

©2017, Toshiya Hatano, Masaaki Mizuochi, Printed in Japan

コラム：2つの分散②

n で割る母分散と $n-1$ で割る標本分散の2つがあるのはなぜでしょうか。本文中では、データの総変動を分解するアプローチにより説明しました。ただし、その際、母平均と標本平均の乖離である標準誤差が σ^2/n となることを既知として扱い、その証明は示しませんでした。以下では、母分散と標本分散について本文中と異なるアプローチで説明を行い、その過程で標準誤差が σ^2/n であることも示しましょう。

いま、母集団の平均（母平均）が μ 、分散（母分散）が σ^2 （標準偏差が σ ）であるとします。この母集団から n の数だけの標本をランダムに抽出するものとし、抽出された標本を x_i (i は $1 \sim n$)と表記します。

まず、分散を計算するための準備として、標本の偏差平方和 $\sum_{i=1}^n (x_i - \bar{x})^2$ について考えてみましょう。 μ を利用して偏差平方を2つに分解し、展開した上で Σ 記号の一部を演算すると、(1)式が得られます。

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n [(x_i - \mu) + (\mu - \bar{x})]^2 \\
 &= \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2 \\
 &= \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu) \cdot n(\bar{x} - \mu) + n(\bar{x} - \mu)^2 \\
 &= \sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2 \tag{1}
 \end{aligned}$$

次に、(1)式の第1項 $\sum_{i=1}^n (x_i - \mu)^2$ について考えます。 $(x_i - \mu)^2$ は各標本値と母平均の差の2乗です。つまり、期待値としては母分散 σ^2 に等しくなるはずですが、これを n 個足し合わせるのですから、第1項は、

$$\sum_{i=1}^n (x_i - \mu)^2 = n\sigma^2 \tag{2}$$

となります。第2項 $n(\bar{x} - \mu)^2$ のうち $(\bar{x} - \mu)^2$ は標本平均と母平均の差の2乗です。標本平均 \bar{x} を、 x_i を用いて書き直すと、 $(\bar{x} - \mu)^2$ は、

$$(\bar{x} - \mu)^2 = \left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu) \right]^2 = \frac{1}{n^2} \left[\sum_{i=1}^n (x_i - \mu) \right]^2$$

となります。標本 x_i は平均的には μ に等しく（つまり、 $x_i - \mu$ の期待値は 0 となり）、標準偏差は σ ですから、標準偏差の合成法則（ $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ ）と $\sigma_i = \sigma$ に注意すると、

$$(\bar{x} - \mu)^2 = \frac{1}{n^2} \left[\sqrt{\sum_{i=1}^n \sigma_i^2} \right]^2 = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

が得られ、この式から、

$$n(\bar{x} - \mu)^2 = \sigma^2 \quad (3)$$

であることがわかります。(2)式と(3)式を(1)式に代入して整理すると、

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

が得られます。この式を変形すると、

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

となり、標本と標本平均による偏差平方和 $\sum_{i=1}^n (x_i - \bar{x})^2$ を $n-1$ で割ると、母分散 σ^2 をうまく推定できることがわかります。