

数値の有効桁数について

データの平均や標準偏差を書くときに、どの桁まで表示しているであろうか。ここでの考え方は、例えば5段階評価などにも応用できる。今、 n 個の観測データがあるものとする。このとき、個体 i のデータが以下のような形になっているものとする。

$$X_i = x_i + \varepsilon_i \quad (\text{データ} = \text{読み取った数値} \pm \text{読み取り誤差})$$

ここで簡単のため、読み取り誤差は ± 0.5 の範囲にあると考える。また読み取り誤差の分布としてはどこに数値があるか分からないから、この区間の一様分布である。この分布の平均は0であり、分散は $1/12=0.0833$ 、標準偏差は0.2887、個体により独立である。

$$E[\varepsilon_i] = E[\varepsilon] = 0, \quad V[\varepsilon_i] = V[\varepsilon] = \int_{-0.5}^{0.5} x^2 dx = \frac{1}{12} \cong 0.0833$$

これらの n 個のデータの平均は以下である。

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \varepsilon_i$$

データの標本平均の平均及び、標本平均の分散は、以下である。

$$E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i + \varepsilon_i)\right] = \frac{1}{n} \sum_{i=1}^n x_i$$
$$V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = V\left[\frac{1}{n} \sum_{i=1}^n (X_i - x_i)\right] = V\left[\frac{1}{n} \sum_{i=1}^n \varepsilon_i\right] = \frac{V[\varepsilon]}{n}$$

これらの関係を平均値と測定の標準偏差を用いて形式的に書くと

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n x_i \pm \frac{\sqrt{V[\varepsilon]}}{\sqrt{n}} \quad (1)$$

データの標本分散の平均は、誤差が十分小さいとして以下である。

$$E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = E\left[\frac{1}{n} \sum_{i=1}^n \{(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})\}^2\right]$$
$$= E\left[\frac{1}{n} \sum_{i=1}^n \{(x_i - \bar{x})^2 + (\varepsilon_i - \bar{\varepsilon})^2\}\right] = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + V[\varepsilon] \approx \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

データの標本分散の分散を計算するには誤差が十分小さいとして以下の関係を利用する。

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right]$$
$$= \frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})]^2 - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 - V[\varepsilon]$$
$$= \frac{2}{n} \sum_{i=1}^n (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) + \frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - V[\varepsilon] \approx \frac{2}{n} \sum_{i=1}^n (x_i - \bar{x})\varepsilon_i$$

この関係を用いて、標本分散の分散は、誤差が十分小さいとして以下である。

$$V\left[\frac{1}{n}\sum_{i=1}^n(X_i - \bar{X})^2\right] \approx \frac{4}{n^2}V\left[\sum_{i=1}^n(x_i - \bar{x})\varepsilon_i\right] = \frac{4}{n^2}\sum_{i=1}^n(x_i - \bar{x})^2V[\varepsilon] = \frac{4}{n}S^2V[\varepsilon]$$

これらの関係を分散と分散の標準偏差を用いて形式的に書くと、以下のようになる。

$$\frac{1}{n}\sum_{i=1}^n(X_i - \bar{X})^2 = S^2 \pm \frac{2S\sqrt{V[\varepsilon]}}{\sqrt{n}}$$

また、これから標本標準偏差は以下のようになる。

$$\sqrt{\frac{1}{n}\sum_{i=1}^n(X_i - \bar{X})^2} = \sqrt{S^2 \pm \frac{2S\sqrt{V[\varepsilon]}}{\sqrt{n}}} = S\sqrt{1 \pm \frac{2\sqrt{V[\varepsilon]}}{S\sqrt{n}}} \approx S \pm \frac{\sqrt{V[\varepsilon]}}{\sqrt{n}} \quad (2)$$

(1), (2) 式より、測定標準偏差（測定の誤差）は平均値、標準偏差とも $0.2887/\sqrt{n}$ となる。

具体的には、測定標準偏差は以下のようになる。

$n=10$ のとき、測定標準偏差は、0.0913

$n=50$ のとき、測定標準偏差は、0.0408

$n=100$ のとき、測定標準偏差は、0.0289

$n=1000$ のとき、測定標準偏差は、0.0091

これを見ると、用心のため誤差を測定標準偏差の 2 倍まで取るとして、やはり $n=100$ くらいまでは実測値から 1 桁低い値で表記するのがよく、それ以上数百程度になると 2 桁下げるのがよいように思われる。