2020/10/17 畠山

平均値と有効数字

問題: 有効数字 2 桁で、20 という値が 5 回連続で計測された(20,20,20,20,20)場合の平均値はいくつか?

回答 A 有効数字の計算規則を使う

次の有名な規則に従って計算をしてみる。

和と差: 測定値の最小の桁の位に合わせる
積と商: 数値の有効数字の桁数の一番小さいものに合わせる 規則 1

$$\left(\frac{20+20+20+20+20}{5}\right) = \frac{100}{5} = 20.0 \qquad \text{数式 1}$$

<u>つまり、平均値は 20.0</u>

回答 A' 有効数字の計算規則を使う(その 2)

数式 1 において、先に割り算をしても数学的には間違いではないので

$$\left(\frac{20+20+20+20+20}{5}\right) = 4+4+4+4+4 = 20 \qquad \text{数式 2}$$

<u>つまり、平均値は 20</u>

回答 B 不等式を使う

20 というのは、有効数字の意味を考えると、20±0.5 (数式 3)程度の不確定性を持つ。

$$19.5 \le 20 < 20.5 \qquad \text{数式 3}$$

この不等式を各計測値について足し算すると

$$\begin{gathered} 19.5 \le 20 < 20.5 \\ + \\ 19.5 \le 20 < 20.5 \\ + \\ 19.5 \le 20 < 20.5 \\ + \\ 19.5 \le 20 < 20.5 \end{gathered} \qquad \text{数式 4}$$

$$+$$
$$19.5 \leq 20 < 20.5$$

つまり

$$19.5 \times 5 \leq (20 \times 5) < 20.5 \times 5 \qquad \text{数式 } 5$$

5 点のデータの平均値を考えると

$$(19.5 \times 5)/5 \leq \text{平均値} < (20.5 \times 5)/5 \qquad \text{数式 } 6$$

$$19.5 \leq \text{平均値} < 20.5 \qquad \text{数式 } 7$$

<u>つまり、平均値は 20</u>

なぜ、計算法によって有効数字が変わってしまうか?

回答 A,A',B では、異なる計算結果が得られてしまった。その理由は、有効数字という概念のあいまいさに由来している。有効数字は便宜的に作られた概念であり、必ずしも数学・統計学的に正しいものではない (統計的に正当なアプローチは後述する)。

回答 A では仮定した「規則 1」の運用に問題がある。この規則は「このように運用すると、そこそこの精度が得られますよ」という程度の目安に過ぎず、常に有効性を保証している訳ではない。特に、今回のように四則演算に何度も繰り返してしまうと、計算結果の差が生まれやすい。

本手法を濫用すると、計測結果の微妙な違いで有効数字の桁数が変わってしまうのも問題である。

計測値が(20,20,20,20,20)だった場合は

$$\left(\frac{20+20+20+20+20}{5}\right)=\frac{100}{5}=20.0 \qquad \text{数式 8}$$

計測値が(20,20,20,20,19)だった場合は

$$\left(\frac{20+20+20+20+19}{5}\right)=\frac{99}{5}=20 \qquad \text{数式 9}$$

計測値が(50,50)だった場合は

$$\left(\frac{50+50}{2}\right)=\frac{100}{2}=50.0 \qquad \text{数式 10}$$

計測値が(20,20)だった場合は

$$\left(\frac{20+20}{2}\right)=\frac{20}{2}=20 \qquad \text{数式 11}$$

20 という計測値を 5000 回得たとすると

$$\left(\frac{20\times5000}{5000}\right)=\frac{100000}{5000}=20.000 \qquad \text{数式 12}$$

±0.5 の精度で計測していることを鑑みると、上記の結果は現実に即していないようにも感じられる。

一方で回答 B は不等式を考えている。±0.5 の精度で計測しているので、

$$19.5 \leq \text{平均値} < 20.5$$

という結果は妥当とも思える。

しかし、5 回も計測して 20 という値が何度も出たのだから、もとの計測値(20±0.5)よりは、もう少し精度が高くなっても良いのでは、という気もする。

回答 C 誤差伝搬の法則を使う

回答 A、B(and A')は有効数字をそれぞれ過大、過小評価しすぎている可能性がある。誤差伝搬の式(詳細は Web や書籍等を参照)を使うと、誤差範囲を統計的に正しく評価できる。

x_iを測定値としたときの、平均値$\bar{x}$は

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i \qquad \text{数式 13}$$

x_iが誤差 σ を含むとすると、誤差の平均σ_{ave}^2は誤差伝搬の式より

$$\begin{aligned}\sigma_{\text{ave}}^2 &= \sum_{i}^{N}\left(\frac{\partial \bar{x}}{\partial x_i}\right)^2 \sigma^2 \\ &= \sum_{i}^{N}\left(\frac{1}{n}\right)^2 \sigma^2\end{aligned} \qquad \text{数式 14}$$

つまり

$$\sigma_{\text{ave}} = \frac{1}{\sqrt{n}}\sigma \qquad \text{数式 15}$$

今回はσ=0.5、N=5 なのでσ_{ave} =0.22

まとめると

種別	計算結果	備考
回答 A	20.0±0.05	精度を過大評価
回答 B	20±0.5	精度を過小評価
回答 C	20±0.2	○

参考 C 誤差伝搬の法則を使う(その 2)

誤差伝搬の式を四則演算に適用すると、計測値 n± σ、n'± σ'について、次の式が得られる。

$$
\begin{gathered}
(n \pm \sigma) \pm (n' \pm \sigma') = (n \pm n') \pm \sqrt{\sigma + \sigma'} \\
(n \pm \sigma) \pm (n' \pm \sigma') \pm (n'' \pm \sigma'') = (n \pm n' \pm n'') \pm \sqrt{\sigma + \sigma' + \sigma''} \\
\cdots
\end{gathered}
\qquad \text{数式 16}
$$

$$
(n \pm \sigma) \times or \div (n' \pm \sigma') = (n \times or \div n') \pm (n \times or \div n') \sqrt{\left(\frac{\sigma}{n}\right)^2 + \left(\frac{\sigma'}{n'}\right)^2} \qquad \text{数式 17}
$$

数式 16、数式 17 をもとに計算をしてみる。計測値の和は

$$
\begin{gathered}
(20 \pm 0.5) + (20 \pm 0.5) + \cdots + (20 \pm 0.5) = \\
= 20 + 20 + \cdots + 20 \pm \sqrt{0.5 + \cdots + 0.5} \\
= 100 \pm \sqrt{2.5}
\end{gathered}
\qquad \text{数式 18}
$$

よって、平均値は

$$
\begin{gathered}
\frac{100 \pm \sqrt{2.5}}{5} = \frac{100}{5} \pm \frac{100}{5}\sqrt{\left(\frac{2.5}{100}\right)^2} \\
\underline{= 20 \pm 0.3}
\end{gathered}
\qquad \text{数式 19}
$$

あるいは、 A'と同様、先に割り算をしてみる。

$$
\begin{gathered}
\frac{(20 \pm 0.5)}{5} = 4 \pm \frac{20}{5}\sqrt{\left(\frac{0.5}{20}\right)^2} \\
= 4 \pm 0.1
\end{gathered}
\qquad \text{数式 20}
$$

データの和は

$$
\begin{gathered}
= 4 + 4 + \cdots + 4 \pm \sqrt{0.1 + .. + 0.1} \\
= 20 \pm \sqrt{0.5} \\
\underline{= 20 \pm 0.7}
\end{gathered}
\qquad \text{数式 21}
$$

誤差伝搬の式はテイラー展開(1 次)を含む。そのため(?)、上記のように誤差伝搬の式を分割して使っていくと「誤差計算の誤差」が大きくなってしまう模様である。数式 14 のように、平均の式そのものに対して誤差伝搬の式を適用した方が良いようである 。