

群衆の多様性が大きければ大きいほど、予測値の誤差は小さくなる

集合知

↓ 群衆の予測値の分散(ばらつき) = 個々の予測の分散 - 予測の多様性 ↑

$$(C - X)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - X)^2 - \frac{1}{n} \sum_{i=1}^n (x_i - C)^2$$

ここで、 C : 群衆の予測値 X : 真値 x_i : 個々の予測値

式の導出

$$\begin{aligned} (C - X)^2 &= C^2 - 2CX + X^2 = 2C^2 - C^2 - 2CX + X^2 + \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i=1}^n x_i^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2CX + X^2 - \frac{1}{n} \sum_{i=1}^n x_i^2 + 2C^2 - C^2 \end{aligned}$$

$$C = \frac{1}{n} \sum_{i=1}^n x_i \text{ とおくと}$$

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_iX + X^2) - \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_iC + C^2) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - X)^2 - \frac{1}{n} \sum_{i=1}^n (x_i - C)^2 \end{aligned}$$

母集団
元のデータ

母集団の平均値 μ (未知)から、 n 個サンプリングして標本とする。
その標本平均の分布より母集団の μ を推定する

標本1

標本2

標本3

標本k

n 個 ← 多様性

サンプルサイズ n を大きくすると
標準誤差 σ 小 → 推定値の信頼性向上

標本平均値の分布

標本平均値の標準偏差

→ 標準誤差 $\sigma = \sigma / \sqrt{n}$

