平均偏差

大谷（おおたに）　茂（しげる）

§1. はじめに

授業で標準偏差を導入する手順は

(1)　データ $\{x_k\}$ に対し，平均値 m を確認する。

(2)　平均からのずれを計算する。

偏差　$\Delta x_k = x_k - m$

(3)　データの散らばりを表す量を定義したい。
偏差をそのまま平均するとどうなるか。

(4)　$\pm$ で打ち消し合わないように 2 乗して平均する。

分散　$V = (\text{偏差})^2$ の平均

(5)　次元を戻す。**標準偏差**　$\sigma = \sqrt{V}$

賢明な生徒は，手順(4)(5)を不自然に感じる。
($\pm$ を消すだけなら絶対値でいいのに…)
実際，データの散らばりを表す量として，標準偏差の他に「平均偏差」というものがある。

平均偏差　|偏差|の平均

平均偏差は今の学習指導要領では扱わない。しかし，何十年前は，中学校で習った (ような気がする)。
標準偏差と比べて扱いがぞんざいな平均偏差について考察してみた。

§2. 平均偏差の基準値

次の入試問題は平均偏差に関係している。

問題 1 (2001 年　北海道大学・文系)

(1)　$y = |x-1| + |x-2| + |x-3|$ のグラフを描け。

(2)　$g(x) = \sum_{k=1}^{2n+1} |x-k|$ の最小値を求めよ。

解答　(1)　略　(2)

$g(x)$ は連続かつ区分的に 1 次関数。その傾きは

$x \leqq 1$ のとき	$2n+1$
$1 \leqq x \leqq 2$ のとき	$2n-1$
	…
$n \leqq x \leqq n+1$ のとき	-1
$n+1 \leqq x \leqq n+2$ のとき	1
	…
$2n+1 \leqq x$ のとき	$-2n-1$

となるので，$f(x)$ のグラフより最小値は

$$g(n+1) = 2\sum_{k=1}^{n}(n+1-k) = n(n+1)$$

さて，問題 1 の類題を考えよう。

問題 2　$x_1 \leqq x_2 \leqq \cdots \leqq x_n$ とする。

(1)　$f(x) = \dfrac{1}{n}\sum_{k=1}^{n}(x-x_k)^2$ の最小値を与える x の値を求めよ。

(2)　$g(x) = \dfrac{1}{n}\sum_{k=1}^{n}|x-x_k|$ の最小値を与える x の値を求めよ。

解答　(1)　$m = \dfrac{1}{n}\sum_{k=1}^{n} x_k$ とする。

$$f(x) = \frac{1}{n}\sum_{k=1}^{n}(x^2 - 2xx_k + x_k{}^2)$$
$$= x^2 - 2mx + \frac{1}{n}\sum_{k=1}^{n} x_k{}^2$$
$$= (x-m)^2 + \frac{1}{n}\sum_{k=1}^{n} x_k{}^2 - m^2$$

ゆえに $x = m$ のとき最小。

(2)　問題 1 と同様に $g(x)$ のグラフを考えると

n：奇数のとき　$x = x_{\frac{n+1}{2}}$ で最小。

n：偶数のとき　$x_{\frac{n}{2}} \leqq x \leqq x_{\frac{n+2}{2}}$ で最小。

(1)は標準偏差の基準値を平均値にとる理由である。

(2)では最小値を与える x は平均値ではなく中央値になる。だから平均偏差の基準値は中央値の方がよりふさわしいといえる。(残念ながら，平均偏差ではなくなってしまうが)

§3. おわりに

標準偏差の方が重要な最大の理由は正規分布と中心極限定理だろう。これを微分方程式を用いて高校生に証明 (?) させる授業案も機会があれば紹介したい。

(愛知県立旭丘高等学校)