§1.2　代表値

データの要素が実数で表される量であるとします．このとき，データの要素がどのあたりを中心に分布しているのかということがしばしば問題になります．データの要素の分布の中心を示す数量を代表値といいます．つまり，データの代表値が大きい程，そのデータの要素は大きい値を中心に分布しています．

データの代表値として最も普通に使われるのは（相加）平均です．

定義　正の自然数 n に対して n 個の数 $x_1, x_2, x_3, \ldots, x_n$ を要素とするデータの（相加）平均 ((arithmetic) mean) とは次の式の値である：

$$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k \ .$$

通常“平均”というと相加平均を意味しますが，相乗平均といわれる平均もあります．正の自然数 n に対して n 個の正の数 $x_1, x_2, x_3, \ldots, x_n$ を要素とするデータの相乗平均とは次の式の値です：$\sqrt[n]{x_1 x_2 x_3 \cdots x_n}$ ．n 個の正の数 $x_1, x_2, x_3, \ldots, x_n$ を要素とするデータについて，常に相乗平均は相加平均以下になります:

$$\sqrt[n]{x_1 x_2 x_3 \cdots x_n} \leq \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} \ .$$

相加平均より相乗平均の方が有用なことがあります．

各実数 x に対して $\lceil x \rceil$ は x 以上の最小の整数を表します．各実数 x に対して $\lfloor x \rfloor$ は x 以下の最大の整数を表します．

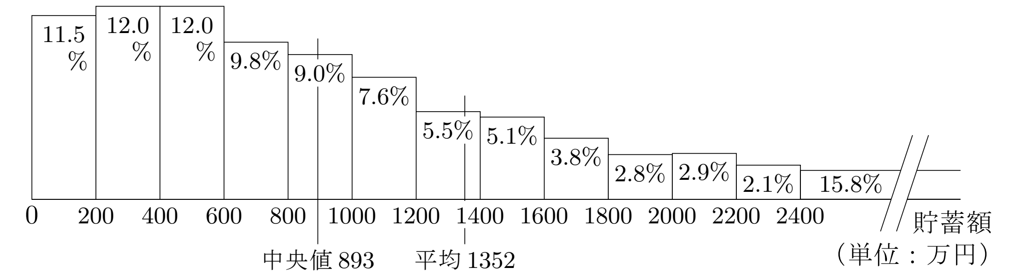
中央値あるいは中位数といわれる代表値もあります．正の自然数 n に対して n 個の実数 $x_1, x_2, x_3, \ldots, x_n$ を要素とするデータの中央値あるいは中位数（メジアン）(median) とは，n 個の値 $x_1, x_2, x_3, \ldots, x_n$ を小さい方から並べ直したときの次の値です：

(1)　n が奇数のときは，中央の値（小さい方から $\left\lceil \frac{n}{2} \right\rceil$ 番目の値）;

(2)　n が偶数のときは，中央の2つの値の相加平均（小さい方から $\frac{n}{2}$ 番目の値と $\left(\frac{n}{2}+1\right)$ 番目の値との相加平均）．

モード（最頻値）(mode) といわれる代表値もあります．これは，データから度数分布表を作ったときに度数が最も大きい階級の階級値です．この値は，度数分布表の作り方（階級の決め方）に依存します．

データの代表値として普通に使われるのは（相加）平均ですが，平均は極端なデータに影響されやすという欠点があります．例えば，総理府統計局が公表した貯蓄動向調査結果によると，1998年末において，勤労者世帯の貯蓄額に対して世帯数の割合をヒストグラムにすると次のようになります．



1世帯の平均は1352万円で，中央値は893万円です．中には大変高額の貯蓄している世帯もあるので，そのような高額貯蓄に引っ張られて平均は大きい方に偏ります．従って，分布の中心を平均の1352万にすることは疑問です．むしろ中央値の893万の方が妥当かもしれません．

このように，度数分布の裾が片方にだけ極端に広がっているとき，平均値は代表値として妥当でないことがあります．

$\boxed{\text{例題 1.2}}$　ある4名の学生の身長を測って次のデータが得られたとする：

$$156\,\mathrm{cm}, \quad 180\,\mathrm{cm}, \quad 162\,\mathrm{cm}, \quad 170\,\mathrm{cm} \ .$$

この4名の学生の身長の平均と中央値とを求める．

4名の学生の身長の平均は

$$\frac{1}{4}(156 + 180 + 162 + 170)\,\mathrm{cm} = 167\,\mathrm{cm} \ .$$

4名の学生の身長の中央値は

$$\frac{162 + 170}{2}\,\mathrm{cm} = 166\,\mathrm{cm} \ .$$

$\boxed{\text{終}}$

$\boxed{\text{問題 1.2}}$　ある試験を受けた4名の学生の点数が以下のようであったとします．

84点，　25点，　76点，　91点．

これら4名の学生の点数の平均と中央値とを求めなさい．