

§ 1.2 代表値

データの要素が実数で表される量であるとする．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る．データの要素の分布の中心を示す数量を代表値という．つまり，データの代表値が大きい程，データの要素の分布の中心は大きい．

データの代表値として最もよく使われるのは**平均** (mean) 厳密にいうと相加平均 (arithmetic mean) である．

[定義] 正の自然数 n に対して n 個の数 $x_1, x_2, x_3, \dots, x_n$ を要素とするデータの (相加) 平均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”というとき相加平均を意味する．また相乗平均 (geometric mean) といわれる平均もある．

[定義] 正の自然数 n に対して n 個の 0 以上の数 $x_1, x_2, x_3, \dots, x_n$ を要素とするデータの相乗平均とは次の式の値である： $\sqrt[n]{x_1 x_2 x_3 \cdots x_n}$ ．

次の定理の証明は省く．

[定理] 正の自然数 n に対して n 個の 0 以上の数 $x_1, x_2, x_3, \dots, x_n$ を要素とするデータについて，相乗平均は相加平均以下である：

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

相加平均より相乗平均の方が有用なことがある．例えば，A 氏と B 氏とは，弓道の同じ 4 つの大会に出場して，A 氏は 4 大会とも 8 本射て的中した本数は 8, 3, 4, 8 であり，B 氏は 4 大会とも 8 本射て的中した本数は 5, 6, 5, 6 であったとする．的中本数の相加平均は，A 氏は 5.75 であり B 氏は 5.5 である．的中本数の相乗平均は，A 氏は約 5.26 であり B 氏は約 5.48 である．的中本数の相加平均を比べると B 氏より A 氏の方が多いが，的中本数の相乗平均を比べると A 氏より B 氏の方が多い．これは，A 氏の的中本数にはばらつきが大きいからである．的中本数のばらつきが小さく安定している方が望ましいと考えるならば相乗平均で比較するほうが合理的である．このように，データの要素の大小のばらつきが小さい方が望ましいときは相加平均より相乗平均で比較するほうが適している．

各実数 x に対して $\lceil x \rceil$ は x 以上の最小の整数を表す²⁾．各実数 x に対して $\lfloor x \rfloor$ は x 以下の最大の整数を表す³⁾．

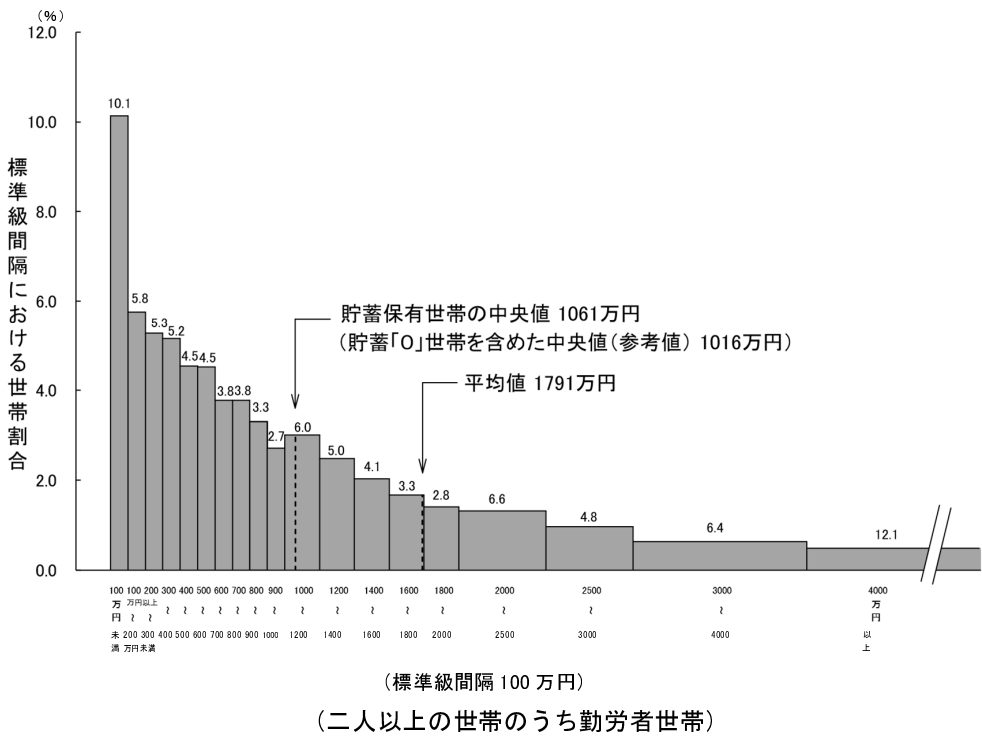
中央値 (メジアン) (median) あるいは中位数といわれる代表値がある．

[定義] 正の自然数 n に対して n 個の実数 $x_1, x_2, x_3, \dots, x_n$ を要素とするデータの中央値あるいは中位数とは， n 個の値 $x_1, x_2, x_3, \dots, x_n$ を小さい方から並べ直したときの次の値である：

- (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 $(\frac{n}{2} + 1)$ 番目の値との相加平均) ．

データの代表値としてもっともよく使われるのは (相加) 平均であるが，平均は極端なデータに影響されやすという欠点がある．

例えば，総理府統計局の資料では，2020 年において，勤労者がいる二人以上の世帯の貯蓄額に対する世帯数の割合をヒストグラムにすると次のようになる．



勤労者がいる二人以上の世帯の貯蓄額について，平均は 1791 万円で，中央値は 1016 万円である．中には大変高額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っている．従って，分布の中心を平均の 1791 万円と考えることは疑問である．むしろ中央値の 1016 万円の方が妥当ではないかと思える．

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る．

[例] 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：

156cm, 180cm, 162cm, 170cm ．

この 4 名の学生の身長の平均は

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

この 4 名の学生の身長の中央値は

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

終

[問1.2] ある試験を受けた 4 名の学生の点数が以下のものであったとする．

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よ．

モード (最頻値) といわれる代表値がある．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ある．この値は，度数分布表の作り方 (階級の決め方) に依存する．

²⁾ 変数 x の関数 $\lceil x \rceil$ は天井関数といわれることがある．

³⁾ 変数 x の関数 $\lfloor x \rfloor$ は床関数といわれることがある．