

「平均値」、「中央値」、「最頻値」について

寺林 暁良

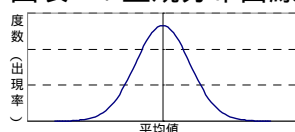
経済統計では、その統計を代表する値（「代表値」）の示し方が重要になる。今回は「代表値」として一般的である「平均値」、「中央値」、「最頻値」を説明する。

「平均値」

「平均値（ここでは、算術平均値）」は、全てのデータを足し合わせ、足し合わせた数で割ることで求められる値である。

平均値は、計算が容易であり、図表1のような正規分布（左右対称のつりがね状の曲線を描くような分布）に従うような統計では利用頻度が高い代表値である。しかし、平均値には、極端な値（外れ値、異常値）の影響を強く受けるという欠点がある。例えば、08年「国民生活基礎調査」の平均所得は556.2万円となっている（図表2）。しかし、年収2,000万円以上には極端に大きな額が含まれ、それらが平均値を大きく引き上げている可能性がある。その場合、実勢よりも所得水準が高めになる点に注意すべきである。

図表1．正規分布曲線



「中央値」

「中央値（中位数、メジアン）」とは、全てのデータを小さい順に並べた時に真ん中に来る値のことである。データが偶数の場合は真ん中2つのデータの平均値が中央値となる。

右に歪んだ分布を示す図表2のように、経済統計は必ずしも正規分布に従うわけではなく、外れ値も含んでいる。右（あるいは左）に歪んだ分布を持ち、外れ値を多く含むような経済統計では、中央値のほうがデータの特性をつかみやすい。

前述の「国民生活基礎調査」における所得金額の中央値は448万円である。平均所得と比較すると100万円以上低い値であるが、所得水準等の場合、中央値のほうが社会の実勢により近いとみなされる場合が多い。例えば、OECDの相対的貧困率の定義は、「等価可処分所得^{（注）}の中央値の半分に満たない世帯員の割合」とされ、中央値が重視されている。

中央値の欠点はデータ量が膨大な場合並べ替えることが困難なことであるが、現在はパソコンの使用によって、並び替え作業も容易になっている。

（注）世帯の可処分所得を世帯人員の平方根で割って調整した所得。

「最頻値」

データの出現率が最大の値を「最頻値（モード）」という。最頻値も中央値と同様、外れ値の影響を受けないという利点があるが、出現率の高い値が複数ある場合は最頻値を一つに決められないことが欠点である。

