

研究員の視点

平均値と実感のズレ

■算術平均の落とし穴

「統計学はよく知らない」という人でも、平均は“なじみ”がある指標の一つではないだろうか。しかし、この平均には様々な種類があり、どのような方法で算出されているのか把握していないと思われ、誤解を生むことがある。

私たちは平均を計算するとき、多くの場合、「算術平均」を使っている。算術平均は、すべてのデータの合計をデータ数で割って求められ、テストの平均点や平均身長などに利用される。

しかし、この算術平均は、極端な値（外れ値）の影響を強く受けるという欠点がある。例えば、総務省の家計調査報告によると、平成22年の二人以上の世帯の平均貯蓄額は1,657万円であった。この金額、思ったより多い、実感と合わない、と感じる人が多いと思われる。

実は、この「平均」が私たちの実感と合うのは、データの分布が下図の折れ線グラフのように、正規分布（左右対称の釣鐘状の曲線を描くような分布）という形に従う場合だけである。

図表にある「貯蓄現在高階級別世帯分布」（棒グラフ）を見ると、全体的に左へ偏った非対称分

布となっており、左右対称の正規分布とはかけ離れた分布であることが一目で分かる。これでは、平均値が多くの世帯の実感とズレが生じるのは無理がない。資産のように上限が決まっていない場合、10.2%を占める貯蓄額4,000万円以上の富裕層が、全体の平均値を押し上げてしまうからだ。

このように正規分布とかけ離れた分布の場合、「中央値」や「最頻値」を見る方が私たちの実感に近づく。

■より実感に近づく中央値と最頻値

「中央値」とは、すべてのデータを小さい順に並べたとき、ちょうど真ん中にくるデータの値のことである。前述「貯蓄現在高」の中央値は995万円と、平均値1,657万円より662万円も低い金額となっている。

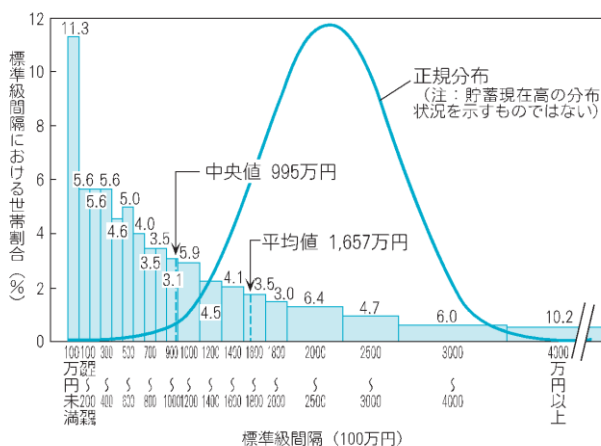
一方、「最頻値」は、基準の数値を一定の範囲で階級分けしたとき、最も多くのデータが集中する階級のことである。前述「貯蓄現在高」では100万円未満の階級となり、全体の11.3%もの世帯が含まれる。この金額であれば、より大多数の人の実感に近づくであろう。

■統計を見るときは元データに近づく

日頃、新聞等で目にする統計数値は、私達が経済や社会の動きを数字で把握する場合、非常に便利なものである。しかし、その数値をただ見たり聞いたりしているだけでは、数値を正確に読みこなすことはできない。一般的に数値には出所の明示がある。インターネット検索等により元データに近づき、数値の裏に潜む真実を推理してみるのもよいのではないだろうか。

(岡本 忠)

貯蓄現在高階級別世帯分布（二人以上の世帯）（平成22年）



資料：総務省「家計調査報告」より引用