

## 標本調査について

岡山 正雄

### 全数調査と標本調査

調査には対象者全員を調査する全数調査と、対象者の中から無作為に選んだ一部を調査する標本調査がある。望ましいのは全数調査だが、国が公表する統計などの場合は、対象者が膨大になり、全数調査をすれば多額の調査費用に加えて、集計に時間がかかりデータの速報性が損なわれる<sup>(注)</sup>。そこで多くの調査では標本調査が行われている。

(注)全数調査である国勢調査の場合、前回調査(2005年)の予算は約 650 億円、最終結果報告書発表に調査開始から約 4 年を要した。

### 標本調査は必ず誤差が生じる

標本調査では一部を調査するため、正しい結果(真の値)との『ずれ』が生じる。このうち、標本調査で導いた値(推定値)と真の値との差を標本誤差という。ここでは完全失業率を例にする。

完全失業率は総務省が行う労働力調査で調べられる。この調査は約 4 万世帯 10 万人を対象としていて 2010 年 2 月の完全失業率(季節調整済)は 4.86% である。

このとき、真の値は 4.73% ~ 4.99% の間に 95% の確率で存在することが示される(図表 1 参照)。なお、4.73% ~ 4.99% は、

推定値と調査対象者数から求められる。

### 調査方法の不備でも「ずれ」が生じる

標本誤差は、無作為に抽出したとしても生じる誤差である。これに加えて、調査方法に不備があるために生じるずれもある。

労働力調査の場合は、層化二段抽出法で対象者を抽出している。この方法はまず全国を 11 地方に分け、地方ごとに全調査区を製造業の多い調査区や学生寮の多い調査区といった特性によって、42 グループに分ける。そして各グループから無作為に 2,912 の調査区を抽出し、さらにその中から調査する約 4 万世帯を抽出する方法である。

このような方法を取る理由は、調査対象全体から無作為に選ぶより、真の値に近い値を求められるからとされる。しかしながら、特性で分けたグループが、正しく調査対象全体の特性を反映せずに、誤った推計値を求める可能性がある。

また、標本調査、全数調査のどちらでも起こることとして、回答者が質問の意味をうまく理解できないために、正確に答えられなかったり、回答を拒否したりする問題がある。その結果、不十分な調査結果から推計しなければならぬことも起こる。

以上のことは調査では避けられない問題であり、調査結果を利用する側はどのようなずれがあるのか把握したうえでデータを見る必要がある。

図表 1: 失業率の標本誤差

