

研究員 の眼

統計データの秘匿

正しく隠すにはどうしたらいいか？

保険研究部 主任研究員 篠原 拓也
(03)3512-1823 tshino@nli-research.co.jp

日本では、様々な統計調査が定期的に行われている。通常、調査結果は、統計表として刊行される。政府が行う統計調査では、全国合計のデータだけではなく、都道府県、更には市区町村レベルでのデータが表示されることも多い。例えば、経済産業省が行っている商業統計では、様々な商業施設の種類ごとに、従業者数、年間商品販売額、商品手持額、売場面積といったデータが、市区町村ごとに表示されている。

そこで、1つ問題が生じる。ある村には、百貨店はなく、スーパーストアが1店だけ営業していたとしよう。このとき、商業統計で、この村の「百貨店、総合スーパー」の欄には、このスーパーストアの情報がそのまま表示されることになってしまう。特定の企業や個人の情報を保護する観点からは、このような表示は避ける必要がある。即ち、統計データを秘匿(ひとく)することが必要となる。

商業統計では、「x」という記号でデータを秘匿している。同統計の利用上の注意によれば、「『x』は1又は2の事業所に関する数値で、これをそのまま掲げると個々の申告者の秘密が漏れるおそれがあるため秘匿した箇所であるが、3以上の事業所に関する数値であっても、前後の関係から秘匿の数値が判明する箇所も同様に秘匿している。」とのことである。実際に、市区町村ごとの商業統計の結果を見ると、多くの市区町村で「x」が並んでいる。

統計データの秘匿にあたっては、どういうデータを秘匿するか、どのように秘匿するか、の2つが悩みどころとなる。

まず、どういうデータを秘匿するか、であるが、データを表示すると、特定の企業や個人の情報が保護できない場合が問題となる。村に1店の場合はもちろんだが、2店の場合でもお互いの店同士で相手の情報がわかってしまうので秘匿の必要がある。村に3店以上何店かあっても、上位の1、2店で売り上げシェアが99%を占めているような場合は、秘匿の必要性があるかもしれない。(ただし、商業

統計では、そこまでは秘匿の対象にしていない。)

次に、どのように秘匿するか、である。1つの方法として、商業統計で行われているように「x」で非表示にしてしまうことが考えられる。ただし、この方法では、統計表の縦横の合計値から、非表示にした数値が逆算できてしまうことがある。そこで、二次秘匿として、3店以上の欄も秘匿して、逆算できないようにする必要がある。この方法では、本来、秘匿の必要性のない欄まで二次秘匿することとなり、せっかく行った統計の結果表示量が減ってしまうという問題がある。

そこで、「x」で非表示にする代わりに、数値を範囲で表示するという方法がある。例えば、秘匿したい欄を「0~10」、二次秘匿の欄を「15~30」などと表示する。しかし、この方法では、あちこちの欄が範囲で表示されることとなり、出来上がった統計表が曖昧なものになってしまう。

範囲での表示ではなく、データを丸めるという方法もある。例えば、データを5単位に丸めて表示する、などとルールを決めておく。7というデータは5に、16というデータは15に丸めることになる。ただし、この方法も、細かいところがわからないという曖昧さはぬぐいきれない。

隣接した市区町村と合算して表示するという方法もある。統計データの表示上で、いわば市区町村の合併をしてしまう方法である。これは、統計表に意味のあるデータを表示するという点ではうまい方法といえるが、当事者の市区町村から見れば簡単には受け入れられないものかもしれない。

このように、データの秘匿というのは、統計の実務では難しい対応が求められる問題である。今後、情報の収集や分析の技術が高度化し、特定の企業や個人の情報の保護がますます求められるようになるものと考えられる。その中で、統計表の作成者には、これまで以上に、どのような秘匿措置をとるべきか、が問われるようになるかもしれない。

統計表を見る側にとっては、秘匿された統計データを目にした際に、どのような意図でこのような秘匿措置がとられているのか、を念頭に置くことも必要ではないかと思われるが、いかがだろうか。