

■ 厚労省の統計不正問題について、(統計は専門ではないが) 高等学校の教科書に載っている学習内容などと絡めて考えてみる。

■ まず報道されたのは、毎月勤労統計調査の調査対象事業者の件数である。従業員数が 500 人以上の事業者は全数調査しなければならないという規定にもかかわらず、東京都では全 1,464 事業者中の約 1/3 の事業者 (491 事業者) を抽出して調査をしていたという。

これによってどのような問題が生じるかという、データに対して信頼度やデータの幅を考える必要が生じるということである。

母平均 m 、母標準偏差 σ の母集団から、大きさ n の標本を無作為に抽出するとき、標本平均 \bar{X} は、 n が大きいとき近似的に正規分布 $N(m, \sigma^2/n)$ に従う (数学 B の基本事項である)。

したがって、この無作為抽出による母平均 m の 95% の信頼区間は $\bar{X} - 1.96 \cdot \sigma / \sqrt{n} \leq m \leq \bar{X} + 1.96 \cdot \sigma / \sqrt{n}$ である。

今回の調査で言えば、(東京都の 500 人以上の事業者については) 抽出したサンプルの賃金平均を出しても、それが真の平均ではなく、95% の信頼度を仮定した場合、 $2 \cdot 1.96 \cdot \sigma / \sqrt{491} \approx 0.1769\sigma$ の幅を母平均に対して考慮する必要のある値である。もしこれが全数調査であれば、このようなデータの幅の考慮は必要なく、100% の信頼度で母平均を計算できたのである。

昨年 11 月の月間現金給与額の平均値は 285,196 円だが、 σ の値がどの程度かは、調査結果をザッと見た限りでは不明だった。 σ を計算しないとは思われないが、公表しない理由があるのだろうか？

似たようなデータを探してみると、「2009 年度能力・仕事別賃金実態調査結果」(公益財団法人 日本生産性本部全国生産性機関) に、一般職 I の賃金平均が 212.3 千円、標準偏差が 51.1 千円とある。これを参考にして、仮に昨年 11 月の東京都の月間現金給与額について $\sigma = 5$ 万円とすれば平均は 95% の信頼度で 8,845 円の幅を考慮すべきデータになる。都の平均値が公表の全体の平均 285,196 円に等しいとして昨年 11 月の月間現金給与額にこの値を適用すれば、都の平均値は 280,774 円 ~ 289,619 円の範囲にある。この幅は半端なく大きい。

■ さらに、復元の問題もある。499 人以下の調査で「1/3 抽出すべきを 1/6 抽出にしてしまった。同じ事業所が 2 つあるような復元をすればよかったが…」などと大臣がしどろもどろに答弁している。

しかし、仮にそのような復元をしても、これによって調査の精度が保たれるということはない。考慮すべきデータの幅は $\sqrt{2}$ 倍に広がってしまうという統計的な知識や認識は、全く欠如している。

彼らが言っている復元は、単に標本平均の帳じりを合わせるだけで、精度への考慮は全くない。大臣はもとより小学生レベルのミスをしている役人たちにも、精度などへの慮りはあろうはずもないか。

■ また、賃金構造基本統計調査においては、調査方法も規定通りではなかったという。訪問調査と言いながら、実際は郵送で済ませていたというのだ。一見何も問題はなさそうに見える。しかし、回収率が低い (2017 年で 72.6%) のは、郵送にしていたからであろう。回収率の低さは精度の低さに直結する。

統計担当職員が削減されてきていて全数調査ができないようなニュアンスの報道や発言も散見されるが、全数調査を抽出調査にし、訪問調査も郵送調査へと仕事量を勝手に減らしたために職員が削減されていったという側面はないのか。卵が先か鶏が先か。

■ 新学習指導要領で、統計教育のウエイトが大幅に上がる (新数 B では、統計を選ばざるを得ないかも!)。背景にはビッグデータなどの有効利用といった数学の社会利用があると思うが、統計の学習においては単に問題解決を図ると言うことのみならず、統計利用上のモラルといったことにまで踏み込んだ教育が必要なのだろう。学習指導要領に、そういったことへの言及はあるだろうか。

高校の統計教育がずっと軽視され続けてきた状況は、センター試験の問題選択率に顕著に表れている。そう言った流れの先に、今回の統計不正問題がある。統計教育について聞かれた文科大臣の答弁 (答弁書を書いている文科省の役人) のお粗末さには、開いた口が塞がらなかった。統計教育の現況を、彼らは全く把握できていない。

『統計でウソをつく法』(ダレル・ハフ著) に統計のウソを見破る方法として「統計の出所」「調査方法」「問題すり替え」など検証すべき 5 項目が列挙されていると、朝日新聞 (2019.2.4) にある。嘘対応が必要な学問である統計とは、つくづく厄介なツールである。