

統計の信頼回復に民間の力を

デジタルデータの有効活用は、 企業の統計報告者負担を軽減する

厚生労働省の毎月勤労統計で不適切な調査が起きた背景には、デジタルデータを省庁に直接渡せないことによる企業の統計報告者負担の重さがある。統計作成をスイスの消費者物価統計のようにデジタル化することで企業側の作業負担が減り、統計作成に要する人件費も節約できる。社会のデジタル化が進む現在、政府より民間に統計作成のアドバンテージがあり、今回の不正は、民間の持つデジタルデータとノウハウを有効に活用するという視点の重要性を示唆している。



東京大学大学院
経済学研究科
教授 渡辺 努

政府統計への信頼が 大きく失墜

厚生労働省が作成している毎月勤労統計の調査が2004年以降、不適切に行われていたことが発覚し、わが国の政府統計への信頼が大きく揺らいでいる。起きたことは単純だ。東京には常用雇用者500人以上の規模の事業所が約1500あるが、このすべてから賃金統計の基

なるデータとして、「現金給与総額」「労働者総数」「総労働時間」を収集するというルールになっていた。しかし、厚労省が実際にデータを収集していたのはその3分の1の約500だった。これは明らかに法令違反だ。ただ、これだけであればまだ対処のしようがあった。未調査の1000の事業所の賃金は調査を実施した500の事業所の賃金と同じと仮定して集計を進

めることは可能だからだ。しかし実際には、厚労省は1000の事業所があたかも世の中に存在しないかのような処理をし、これが公表数値を大きくゆがめた。図表は、厚労省が不正発覚後に公表した再集計の数値を基に、「きまって支給する給与」の14年の値について、再集計と旧集計を比較したものである。500(999人規模の事業所では、

賃金が全体として1・7%、女性に限定すると6%も過小評価されていたことがわかる。全国・全事業規模の平均で見ても賃金水準が0・7%程度、実際よりも低い。こうした過小評価が雇用保険などの過小給付という実害につながった。過小給付は約2000万人に及び、金額にして500億円超といわれている。実は、政府は統計の精度向上

を重要な政策課題と位置付け、17年に官房長官を議長とする統計改革推進会議を設置して検討を重ねてきた。そこでのテーマは、EBPM (Evidence Based Policy Making) であり、国や

自治体の行うさまざまな施策について、事前と事後の評価を科学的な手法で行うことにより、わが国の政策形成プロセスを変えようということである。科学的な評価を支えるのは統計だ。この認識のもと、GDP統計などさまざまな統計の抜本的な見直しが行われ、実行に移す局面であった。その矢先に今回の不正が発覚してしまった。科学的な政策評価が足もとから崩壊したといっても過言ではない。

統計作成技術の遅れ

今回の不正については、組織的な隠蔽があったか否かも含めて国会で審議が行われている最中であり、真相の解明にはもう少し時間がかかりそうだ。以下では、今回の不正が起きたのはなぜかという点について、筆者が研究者として政府統計を使う中で経験したことや、統計改革推進会議のメンバーとして検討に参画する中で学んだことを踏まえて意見を述べたい。

今回の不正を一言でいえば、1500事業所全部を調べるの

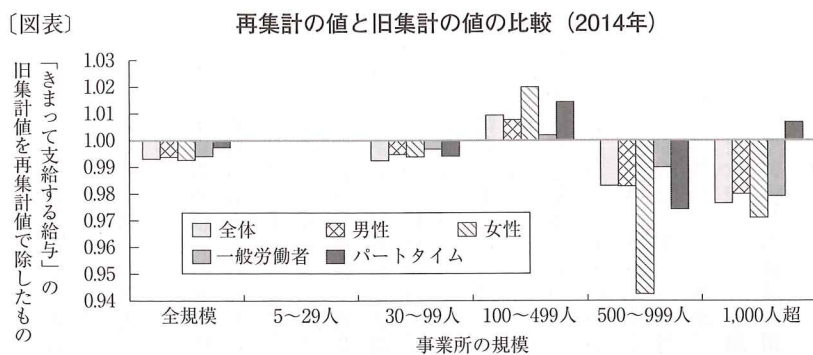
が大変なので手を抜いて間引いたということだ。単純な手抜き以上の悪意があったか否かはわからない。しかし手抜きの意図があったことは間違いない。統計作成の担当者は、なぜ手抜きの誘惑に勝てなかったのか。

筆者は、毎月勤労統計の作成に直接関与したことはないが、調査の現場でどのような作業が行われるかは容易に想像できる。調査は、厚労省が用意する1枚の調査票を企業に埋めてもらうことによって行われる。東京の500人以上の規模の事業所であれば、ほぼ間違いなく、各人の給与額などのデータはデジタル化されている。企業の担当者は、自社のデータベースやクラウドから必要な数字を取り出し、それを基に、その企業の「現金給与額」など厚労省が求める数字を計算する。企業の担当者は、それを調査票に転記し、厚労省に郵送する。調査票を受け取った厚労省は、数値をデータ化したうえで、コンピュータ上で集計作業を行う。おそらくこういうことが毎月行われているのだらう。

一連の作業のどこに問題があるのか。統計報告者の作業負担、おそらく不必要に重い報告者負担を企業の担当者に負わせている点だ。企業の担当者にとって最も嫌な作業は、企業内ではすでにデジタル化されている給与関連の情報を調査票に転記しなくてはならないことではないか。もし企業のコンピュータから厚労省のコンピュータに給与関連のデータをコピーできるのであれば、手間は大幅に軽減できる。実際、企業内の日常業務ではそうやってデータを活用しているはずだ。しかし、厚労省に情報を渡そうとすると、紙の調査票に転記するという余計な手間がかかってしまう。企業はこうした負担を嫌い、回答を渋る。

企業がすでにデータをデジタル化しているにもかかわらず、直接省庁に渡せないという非効率な状態は、実はさまざまな統計で起きている。例えば、筆者が頻繁に使用する消費者物価統計の場合、店舗側ではどの商品がいくらで何個売れたかをデータとして保有している。少し先進的な店舗であれば、その商品

再集計の値と旧集計の値の比較 (2014年)



〔図表〕
旧集計値を再集計値で除したものの「きまって支給する給与」の

の購入者がどういう属性の人で、過去にどういう購買履歴があるかもデータとして保有し、マーケティングや商品開発に活用している。

しかし消費者物価統計を作成する総務省統計局が、こうしたデジタルデータを利用することはほとんどない。総務省の手法はICT（情報通信技術）普及前と同じであり、調査員が店舗を訪問し、店頭の価格を目で見て、それを手もとの調査票に記入する。筆者の手もとには1950年当時の調査票があるが、現在の調査票と基本的に中身は同じだ。例えば、胃腸薬の項目を見ると、当時の「太田胃散缶入り81g」が今は「第一三共胃腸薬細粒」に替わっているが、違いはそれだけで、今も昔も調査員が店頭で胃腸薬の値段を調べるスタイルだ。

今回の不正の原因として、統計作成に携わる人員の削減が行われてきたことを指摘する声が少なくない。財政再建のために人員削減が行われており、そのために統計作成の現場が疲弊していることは事実である。しか

し、少ない人数でやれないのかといえはけつしてそうではない。企業が持つデジタル化されたデータを省庁がそのまま受け取ることができれば、大幅な省力化が可能だ。筆者は、今回の不正の底流には、昔ながらの調査手法を変えようという省庁の姿勢があると考えている。

統計作成のデジタル化

それでは、統計作成のデジタル化は本当に実現可能なのか。筆者は消費者物価統計の推計手法に関する研究の一環として、各国で採られている手法について統計部署の担当者意見交換する機会が多い。筆者の印象では、欧州の国々はデジタル化に熱心に取り組んでおり、実用化に成功している例も少なくない。

例えば、スイスでは連邦統計局が大手スーパーの協力を得て、商品価格のデータを提供してもらい、それを用いて消費者物価指数を作り始めている。スイスも従来は調査員が店頭足を運んで商品の価格を調べていたが、その手間が大幅に削減された。

店舗側にも、日中の忙しいときに調査員の相手をしなくて済むというメリットがある。スイス連邦統計局では、こうして浮いた人手をデジタル化の難しい分野に回したり、オンライン商取引などこれまで手付かずの分野の統計開発に充てたりして、統計の充実を進めている。同様の取り組みはスウェーデン、オランダなどでも進められている。

スイスなどの取組みで興味深いのは、データ取得の方法の変更が統計の連続性に与える影響を極力なくそうとしていることだ。例えば、スイス連邦統計局が現在入手しているデジタルデータに含まれている商品の数は以前の何十倍にもなる。逆に言えば、調査員による調査のときには世の中にある商品の数%しか観察することができなかった。せっかく全商品に関する情報を入手したのだから、この際、全商品の情報を用いる物価統計に刷新すればよいと思うのだが、実際には、調査員の時代と同じ商品のデータしか使わずその他は捨てることにより、過去の統計との連続性を維持しようとし

ている。そこまでして連続性にこだわるべきか否かは議論の分かれるところかもしれないが、少なくとも、デジタル化を進めると連続性が犠牲になるというしばしば聞かれる意見は当たらない。

スイスなどの事例に共通するのは、報告者負担の重さに対する企業の不満が発端となって、統計局との間でデジタル化の話合いが始まるという構図だ。一方、統計局のスタッフの一部には、昔ながらの手法を変えることに強い抵抗があり、この解消にかなりの時間がかかる。筆者の見る限りわが国も似た状況にある。

デジタル化の具体例をもう一つ紹介しよう。筆者の研究室では、日本経済新聞社の協力を得て、13年春から「東大日次物価指数」を配信してきた（16年にナウキャストに移管）。これには物価指数の計測手法に関する研究成果を社会還元するという目的に加え、統計のデジタル化の実証実験という側面もあった。どの店でどの商品がいくらで何個売れたかという情報を、日

経と契約した約300の店舗がその日の夜に日経のサーバーに送り、それをさらに東大のサーバーに転送する。東大が開発したアルゴリズムを用いて物価指数を計算し、それを翌々日にはホームページにアップする。この間、東大側は完全に無人で、誰かがデータに触るのは年に数回起きるシステムの不具合のときだけだ。更新頻度も、総務省の消費者物価指数が月1回なのに対して、同指数は毎日更新される。

当時、同指数が取り扱っていたのはスーパーで扱う品目に限定されており、総務省の消費者物価指数がカバーする品目の約2割だ。それらの品目の物価指数を作成するために総務省とその作業を実際に担当する各県でどれだけのマンパワーが割かれているのかは不明だが、その部分だけでも無人化できればマンパワーを大幅に削減できるはずだ。スイスのように、デジタル化が容易でない分野に人材を回すことにより、省庁全体として統計作成の生産性を向上させることができる。

この話を省庁にすると、民間のデータは高くても手が出ないの

で実現性が低いとすぐ拒否される。確かに東大指数で使っている購買履歴データは、商用利用を前提に値段が付いているので高い。省庁の理屈でいえば、いまはただで取得できている情報に高額な費用を支払うのは筋が通らないということかもしれない。しかし省庁は、近年減少したとはいえ、今なお膨大な数の調査員を抱え、統計作成にかなりの人件費をかけている。また、調査員が日中の忙しいときに店舗を訪問するなど報告者の側にも費用が発生している。データ購入費用と比べるべきは、これらすべての費用の合計だ。消費者物価指数についていえば、筆者の概算では、デジタル化でお釣りがくるはずだ。

統計の民営化に向けて

今回の不正で明らかになったことは、省庁は民間に比べて統計作成にアドバンテージがあるとは言えないということだ。統計の基になるデータを握っている

るのは誰かといえば民間企業だ。そのデータの加工作業も民間企業のほうがはるかに長けている。もちろん昔からこうだったわけではない。昭和あるいはそれ以前の時代であれば、統計の基になるデータはどこにも存在せず、政府が調査員を現場に派遣し、ミクロの情報を収集する以外に手はなかった。また、集めたデータの加工ノウハウも政府が握っていた。あらゆる意味で政府が民間を凌駕していた。しかし、ICTの進展に伴って、政府と民間の立場が逆転している。

筆者はかつて「統計民営化」の可能性を提唱したことがある（注）。発想は単純で、鉄道サービスを提供主体が政府から民間企業へと変わっていったように、統計サービスの提供主体も政府から民間企業に変わるときがくるだろうということだ。

執筆後2年が経ち、事態は当時想定していたよりも大幅に速いペースで進行していると感じる。例えば、投資家は、景気を知りたいときに各国政府が公表するGDPではなく、マーケット

ット社の作成するPMI（購買担当者景気指数）を見るようになっていく。タイムリーで精度も高く、国際比較も可能だからだ。クレジットカードの購買履歴情報から消費の先行きを占う指標が作られるなど、いわゆるオルタナティブデータの普及も予想を上回るペースで進んでいる。

今回の不正は、政府の統計作成技術が劣化していることを強く示唆しており、徹底的な点検と修復が必要だ。ただし、その作業の目的は、何十年か前の統計制度に戻すということではない。民間の持つデジタルデータとノウハウを有効に活用するという視点が欠かせない。

（注）柳川範之・渡辺努「民間による経済統計の革新…統計民営化の可能性も含めて」（『統計』2017年1月号）

わたなべ つとむ
82年東京大学経済学部卒、日本銀行入行。92年ハーバード大学Ph.D. 11年から現職。