

歴史データの作り方と使い方

有本 寛

筆者はこれまで、開発経済学と経済史の狭間で、戦前の日本の歴史データを使った研究を行ってきた。この経験をもとに、日本の歴史データの作り方と使い方を紹介したい。

●歴史データの作り方

定量的な統計分析に使えるような、日本の歴史的な資料は、実はアクセスが比較的容易なものが多い。『府県統計書』や『工場統計表』など、地方自治体や政府が作成した統計は刊行されており、多くの図書館に所蔵されている。一部は国立国会図書館の「近代デジタルライブラリー」からオンラインでも画像として閲覧できる。刊

ある。

資料へのアクセスは比較的容易だが、統計分析ができるような数値データにするには、資料の撮影、印刷、データの入力、精査等を自らの手で行わねばならない。それなりのやる気と根気と元気が求められる。ただし、これらは資金で代替できる。それぞれの作業は定型なので、アルバイトや業者に委託することが可能である。とはいえ、少なくとも一部は入力からチェックまで自ら経験しておくことよいだろう。データを入力する点とて、統計の平均やばらつきに関する「手触り感」を得ることができし、統計のクセや潜むデータの誤りに気づくこともある。

の長期統計系列⁽¹⁾には、全国あるいは都道府県レベルで、国土、人口、経済、社会、文化等の統計が、古いものでは明治初期から公開されている。農業関係では、農林業センサスの累年統計⁽²⁾が有用である。都道府県ごとに土地や人口といった基礎情報から、農商工業統計、はては伝染病患者死亡数といった、さまざまな統計指標を掲載している『府県統計書』も、古いものから数値データ化してホームページで公開している都道府県もある。手入力する前に確認するとよい。ただし、これらの政府統計の多くは、印刷用に体裁を視覚的に整えたExcelファイルそのまま公開しており、統計分析を行うには著しく不便である⁽³⁾。分析に至るまでに、結合されたセルを整理したり、不要な空白を取り除いたり、変数名を整理したり

と、煩雑な作業が待っていることを覚悟されたい。

歴史データには誤りが多い。政府等の公式統計であっても、誤植等がそれなりに散見される。これは例えば、市町村レベルの数値を県レベルで集計してみると、資料の県レベルの数値と異なることなどから発覚する。こうした場合、筆者はできるだけ修正を試みてきた。例えば、市町村レベルのデータの生産量に誤りがあると思われる場合は、作付面積と反収^{たんしゅう}のデータと突き合わせたり、前後の年や周辺の市町村や家計のデータと比べてたりするなどして、修正ができることがある。よくあるのは、数字が隣の位と入れ替わっているケースである。ただし、どうしても修正ができない場合も多い。正しいデータが永遠に得られないという事実、統計データを正確に作ることの歴史的な重みと責任を感じるものである。

データの誤りは、データ（資料）そのものか、入力ミスのいずれかに起因する。誤りの原因がデータにあることを特定するためにも、データ入力は個別に二回行い、それらを照合することで、入力による誤りを潰しておくことよい。二度

手間のようだが、筆者の経験では、事後的な精査の手間暇を考えると、この方が早い。

●歴史データの使い方

歴史データを使うにあたってもっとも考慮すべきは、そもそもなぜ歴史データを使うかである。歴史データは、現代のデータと比べてカバレッジが狭かったり、パネル化が困難であったり、必要な情報が欠けていたり、不利な制約があることが多い。それにも関わらず、歴史データを使うべき必然性は何か、研究や分析の目的や課題に照らして、説得的に説明することが求められる。

歴史データを使うもっとも直接的な理由は、その時代のその対象について知りたいということだろう。この場合、歴史的な事実の解明や記述が研究の目的となり、動機としては歴史研究に近い。日本の歴史研究には、主に定性的なアプローチに基づいた厚い蓄積があり、事実関係はかなり詳細まで明らかにされていることが多い。定量的なアプローチで、それをどのようにに補完し、歴史研究に貢献できるかが、歴史データを使った定量的な研究を位置づける重要なポ

イントだろう。

歴史データを使ういまひとつの理由は、より一般的な動機に基づく課題の適切な対象が、歴史的な事例だったということである。工場レベルの統計を使った筆者自身に日本から朝鮮への直接投資が、地場工場の参入に与えた影響を検証した、李チャンミン氏（福岡県立大学）との共同研究⁽⁴⁾である。

この研究の目的は、技術力や資本力の高い海外からの直接投資（FDI）が、地場企業の参入を阻んでしまっているのではないかという、一般的な懸念を検証することであった。この懸念は特に、発展途上の経済で、投資元と投資先の間で技術力に大きなギャップがあるときに顕著であると考えられる。これを検証するには、投資先の国において、FDIの動向と地場企業の参入・退出がトレースできるデータが必要だが、現代の発展途上国ではそれがなかなか得られない。戦前の日本から植民地朝鮮への直接投資は、これに適した事例だった。日本は第一次世界大戦を期に重工業化が進み、技術も飛躍的に向上したのに対して、朝鮮は農業を中心とした経済であり、研究の

コンテキストとして適切だった。さらに、朝鮮総督府が調査・刊行した『朝鮮工場名簿』から朝鮮の一定規模以上の全工場の情報が得られ、工場の所有者の氏名から所有者が日本人か否か（つまり日本からの直接投資か否か）が判明した。こうして、地域・産業別に、日本から進出した工場と地場工場の数がわかるため、課題の検証が可能となった。

このように書くと、先に研究目的があり、それに合う事例を選定したようにみえるかもしれないが、実際は同時進行である。「こういうデータがある」という知識と、漠然としたさまざまな問題関心を頭に入れて、脳裏で熟慮を重ねると、おのずとデータと研究課題のマッチングはできてくるように思われる。

●おわりに

日本の歴史データは、世界的にみても充実しており、アクセスや利用可能性も低くない。ぜひ有効かつ積極的に利用されることを願いたい。ただし、なぜ歴史データを使うのか、その必然性を研究の目的に照らして意識することが重要である。また、歴史データは、

厳密な仮説検証や因果関係の識別に耐えられないことが多い。統計的な事実を過剰に解釈しない点にも節度が求められるだろう。

（ありもと ゆたか／アジア経済研究所 ミクロ経済分析研究センター）

《注》

(1) <http://www.stat.go.jp/data/chouki/>

(2) <http://www.maff.go.jp/j/tokei/census/afc/past/stats.html>

(3) このような excel ファイルは俗に「ネ申（＝紙）Excel」と呼ばれる（奥村晴彦、二〇一三『「ネ申 Excel 問題」』。 <http://oku.edumie-u.ac.jp/~okumura/SSSS2013.pdf>）

(4) Arimoto, Yutaka, and Lee, Changmin 2014. "Did Japanese Direct Investment in Korea Suppress Indigenous Industrialization in the 1930s? Evidence from county-level factory entry patterns." IDE Discussion Papers, No.450.