数研 AGORA

【統計特集】

- ▶ 社会科で「統計データ」を使う楽しさ
／小原 美紀……1
- ▶ 社会科の授業における統計の学習を通じた社会科学教育とデータサイエンス教育の可能性を考える
／児玉 英靖……5

- ▶ 算数・数学科で学ぶ統計分野の内容
／数研出版数学編集部……9

- ▶ 「歴史総合」の授業はどのように取り組まれてきたのか
／矢部 正明……12

No.82

この用紙は、再生紙を使用しています。

社会科で「統計データ」を使う楽しさ

大阪大学教授
小原 美紀

1. 統計データを手軽に使える時代

人の行動をデータで捉えられないか、と日々奮闘している。とくに、環境や社会政策が人の行動をどう変えて、その変化が政策をさらに動かす様子をデータで追いかけてられないか、数学や統計学を使って時代を生きる人を描写できないか、と考えている。

そんな私にとって統計データはとても大切なものだ。しかしながら、私は、科学的根拠として統計データが最重要だとは思っていない。統計データが示すものは、たくさんある証拠の一つに過ぎない。そもそも、最初に経済学の研究を始めた頃、自分がやっているデータ分析は世の中に受け入れられにくいと感じていた。経済学分野以外の人に批判され続けたのは「データで人は捉えられない」「人の行動は数字では表せられない」だった。経済学分野の人に言われたのは「理論(数理モデル)が示すことに間違いはない。理論に合わないならば、データの問題か、統計学の使い方の問題」だった。「理論に合うデータを探してくれないか」という人もいた。

時代は変わった。統計データがあらゆる場面で取り上げられる。一般紙にも、数値や統計指標が登場する。以前より、統計データを使った説明が社会に受け入れられるようになったと感じる。説明に統計データを入れると信じてもらいやすい時さえある。データが示すのだから間違いはない、とか、データがないから信じられないという人もいる。

そして、この変わった時代に違和感がある。統計データが示すことは決して完全ではない。統計は平

均的な行動を表現するのに長けているが、社会には平均的でない人もいる。統計はときに特異な行動も捉えるが、それは決して社会の代表的な姿ではない。これらを理解して統計データを使う謙虚さがなければ、人の行動を理解しようとする、そして政策の在り方を考えようとする思考プロセスの害にさえなると思う。

それでは、なるべく間違いなく、データから実態を把握するためには何を理解しておく必要があるだろうか。とくに、文系科目である社会科(公民科・地理歴史科)で、生徒は何を学ぶ必要があるだろうか。もちろん統計学(あるいは一般的な統計)の知識が必要である。だが、人や社会のことを考えるために統計データを利用するならば、統計の基礎知識以上に大切なものがある。社会科の知識である。統計データを手軽に使える時代にこそ、高校で習う歴史や地理、倫理、政治・経済、公共の知識が何よりも大切だと感じる。

そもそも、数学の授業で統計学を学んでも、統計データが示すことに興味を持っていない人は多いように思う。数学が得意な人か、我慢して基本を習得できる人(我慢強くてよくわからない数学に耐えられる人)だけが統計データを読めるようになったり、うまく使えるようになったりするわけではない。むしろ、データが示す面白さを最初に味わうことが大切だと思う。必要な基礎知識は後で学べばよい。楽しさを知れば嫌な勉強も我慢できる。高校でデータを見る楽しさを知り、その後、ある人は大学で、別の

人は社会に出てから必要な統計知識を得ればよい。人それぞれ学び方に好みがあると思うが、基礎学力の習得後に応用を学ぶという古典的なやり方でなくてもよいと思う。解析ソフトの発達により、データを眺めることが容易になった今、データが示す世界を楽しんでから、必要性に駆られて統計学を学ぶやり方もありだと思う。そして、社会科は、データを楽しむのに最適な分野だと思う。

2. 社会科の知識がないと統計をうまく使えない

ところで、先日、家族の時間に関するデータを眺めていた。人が何時間働こうとするか、それを家族でどう決めているかは、私が扱う分析テーマの一つである。少し古いが、2016年の「社会生活基本調査」(総務省)をまとめた『令和2年版 男女共同参画白書』(内閣府)によると、30歳代男性の仕事等時間は491分、家事・育児・介護(家事関連)時間は44分前後で、30歳代女性の仕事等時間は249分、家事関連時間は273分となっている。ここで、家事関連時間とは家事、育児、介護・看護、買い物時間の合計で、仕事等時間は通勤・通学、仕事、学習時間の合計である。それぞれ、週全体の時間を7で割った一日当たりの平均時間として求められている。

この統計データから、男性は女性より「市場労働時間」が長く、女性は男性より「家事労働時間」が長いことがわかる。この傾向は30歳代以外の年齢層で見ても同じである。ふと気になった。いつからだろう？これが日本の伝統的な姿なのだろうか。

みんなに持たれている印象ではなく、データで追いかけてみたい。調べてみると、「国民生活時間調査報告書」(NHK放送文化研究所)を見つけた。戦前の生活時間構成(1941年)が紹介されている。以下はその一部抜粋である。

表 戦前の生活時間構成(1941年)(時間・分)

工場労働者	従業	家事	小売業者	従業	家事
男31-45歳	9.50	0.13	男31-45歳	9.38	0.14
女31-45歳	9.26	2.01	女31-45歳	5.08	5.28

この統計データから、戦前も、男性に比べて女性の家事時間が長かったことがわかる。同じタイプの男女で比べても、男女それぞれで従業と家事を比べても同じ示唆を得る。

もう少し掘り下げて、先に見た2016年の様子を、戦前と比べてみたい。ここで困った。どれと比較す

ればよいのだろうか。数値比較の原則は「見たいもの以外の条件は一定に(コントロール)する」である。Aという特徴を炙り出したいならば、A以外の特徴は同じにしてAを比較する。たとえば、同じ性別の同じ年齢で(性別と年齢をコントロールして)、30歳代の女性で家事時間を比較する。

そもそも年齢カテゴリーが異なるので同じ年齢グループでの比較ができない。が、それは大きな問題でないとする。つまり、2016年の30歳代の平均値は45歳まで年齢を広げたとしても大きく変わらないとする(実際、2016年のデータでは男女ともに30歳代と40歳代で平均値はほとんど変わらないから問題なさそうである)。そのうえで、1941年の31-45歳の女性と2016年の30歳代(45歳まで入っていると)の女性を比べてよいのか。75年前とは「30歳代」の社会的立場が違う。取り巻く環境が違う。決めている行動も違う。経済学の言葉で言えば、経済主体が直面している予算制約も時間制約も、最適化する行動も違う。グループ比較が、知りたい比較にならない可能性がある。比較はできても、やってはいけない。

つまり、数字を正しく追いかけていても、得られる答えは正しくない。近年は、より正確で精度の高い比較のための統計値を求めるなど精緻な分析が行われるが、緻密に分析したところで、そもそも比較対象が正しくないのだから意味がない。

ここでは労働時間という身近な例を挙げたので、比較の問題に気づいた。でも、もう少し社会背景がわかりにくい統計データだったらどうだろう。お決まりの「他の条件を一定として注目するものの変化を見る」では、おかしな比較をしてしまう可能性がある。それに気づかずに、出てきた数値の変化について、統計的に有意な差があるといった説明が追加されれば、さらに厄介な問題となり得る。「統計的に有意な」という表現は何となく重要そうな印象を与えるからだ。しまいには、「昔と今では女性の行動がこんなに大きく変わってしまった」という言葉だけが独り歩きしたりする。統計が理解されずに言葉だけが誤って広がるのは珍しくない。誤った解釈が統計エビデンスとして政策議論に使われれば問題は深刻になる。

時間データの話に戻ろう。どう比較しようかと思っていたところ、比較経済史がご専門の斎藤修先生の論文に次の記述を見つけた。「図表2は、NHK調査に基づき、戦前における平日の労働・家事・余暇

時間のパターンを自営業世帯(小売業、農家)と賃金・俸給生活者世帯(俸給生活者、工場労務者)についてまとめた。次に、図表3は総務庁統計局の1991年度『社会生活基本調査報告』から同様の時間統計を示している。ただ、自営業世帯から勤労者世帯へという歴史的位相の変化を考慮して、世帯の累計は共稼夫婦世帯と専業主婦世帯(夫有業、妻無業)の比較となっている(斎藤(2006);11頁)。戦前における調査とは、先に紹介した『国民生活時間調査』の1941/42年調査を指す。

これを読んで当たり前のことに気づいた。1941年の31-45歳の多くは既婚者だろうから、1991年の有配偶世帯と比較した方がよい。年齢層だけをそろえて比較してはいけない。戦前の工場労務者は農業や小売業といった自営業部門とは(自由に労働時間が選べなかったという意味で)大きく異なるから、勤労世帯の典型として取り上げて、1991年の勤労世帯の平均値と比較すればよい。そして、戦前の自営業世帯(小売業や農業世帯)と現代の勤労世帯を比較すれば、日本人の典型的な働き方の変化(日本における労働の変化)を捉えることができる!

世帯形成の歴史的変化がわかっている人には当たり前のことだろう。でも、知識がなければこのような比較はできない。各時代の、社会やそこに生きる人や家族の知識があるからこそ、比較可能なたくさん組み合わせの中から有意義な統計を取り出し、比較できるのだ。社会的背景の知識がなければ、統計データを引っ張ってきてもうまく使えない。

3. 生徒に統計データをどう使わせるか

言うまでもなく、社会的背景の理解には、歴史や地理、倫理、政治・経済、公共を学ぶことが役立つ。そして先に見た通り、社会的背景を理解できれば統計をうまく使える。せっかくだから、社会的背景を学んだ時に、統計データで学びを補足できればもっとよいに違いない。

たとえば、歴史を学ぶ際に、登場するプレーヤー(経済主体:個人だけでなく村や組織、様々な団体を含む)の行動を取り出し、関連データを探し、そのデータを読み取ってみたい。100年前に生きた人の白米と粟・稗の消費量のデータを見たら、生活をイメージしやすくなるのではないか。地域別に比較したら発見があるのではないか。異国の村の洪水記録をまとめたら、はるか遠くの人の生活の大変さを

感じられるのではないか。選挙について学んだら投票率のデータを見て、投票率はなぜこんなに下がっているのかと不思議に思うのではないか。このような作業が、社会を考える力となるだろう。

データを眺めたら、データが言おうとしていることを読み取り、それを人に説明してもらいたい。説明しろと言われてたら、なるべくわかりやすく見せたいと思うだろう。図表をどう作ったらよいか、言いたいことをわかりやすく示す指標はないか、と探すようになる。数学で学んだ平均値と中央値のどちらを使おうか、ばらつきを捉える分散はどう計算するのか、二つの事象の関係を捉える共分散や相関係数、格差を捉えるジニ係数を示せないかなど自分が見せたいものを生徒自ら探すようになれば最高だ。社会科で数学も学べる。

ここで大切なのは、この一連の作業において正解を求めないことだと思う。先生は生徒に面白いかどうかは伝えても、答えは与えなくてよい。答えがないから生徒は考えるし、自由な発想で意見を言える。最終的にわからなければ、もっと勉強したいと思うかもしれない。大学へ行く、あるいは社会に出て学ぶモチベーションとなる。

4. 何に注意したらよいか

データを利用する際には、いくつか注意したいことがある。第一に、出典を確認したい。最近はインターネット上の情報を探ることが多い。それらの中には信ぴょう性のない情報もある。信頼できる出典の基準はないが、たとえば政府統計など多大な時間と資金が投入されて作られたデータは使いやすい。収集工程が明確に示されているからだ。誤りがあった時に訂正が公表されるので、使い方の間違いに気づきやすい。

第二に、調査対象の選択方法(標本の抽出方法)や回収方法を確認したい。統計データの多くは全数調査(特徴を知りたいと考える集団全員を対象に行った調査)に基づいたものではない。標本調査(全体から一部を選んで行われる調査)に基づいたデータである場合、その抽出方法や回収方法を確かめておく必要がある。無作為抽出(あるいは層化抽出)でない場合には、特定の特徴が過度に(過少に)表れてしまう可能性を理解しておきたい。最近は郵送で回収することはなくなったが、元にしていないデータがウェブ調査に基づく場合には、どこでも誰でも回答でき

そうに見えて、高齢層は捕捉されにくいとか、平日の調査は就業者が捕捉されにくい(ので特定の特徴が表れてしまう)といった具合に、調査方法により回答が偏る可能性を理解しておきたい。

第三に、指標や変数の作られ方を確認したい。同じ名前の指標や変数であっても、どの情報に基づいて作られているかが異なる場合がある。たとえば、よく使われる「収入」や「所得」といった変数について、課税前なのか課税後なのか、勤労収入だけなのか非労働収入も入っているのか、個人ベースなのか世帯ベースなのか等により、示されることは異なり得る。格差を捉える「ジニ係数」についても、課税前収入か課税後収入のどちらで計測するかで異なる動きをする。世帯収入か個人収入かでも示唆は異なり得る。個人データではなく、地域グループなどグループ平均値を使って計算したジニ係数は格差を小さく見せる傾向を持つ。計測方法は国によって異なる場合が多いので、ジニ係数を国際比較する時にはさらに注意が必要となる。

第四に、生徒が自ら収集したデータ(インタビュー調査やアンケート調査の回答)を用いて何かを示そうとする場合には、上記3点を明記すること(出典、調査方法や調査対象、変数や指標の作成方法)を覚えておきたい。

最後に、重要な統計値や指標については、出てきた時に、面倒でも定義や指標の特徴を確認しておきたい(数学科の教員と連携すればさらにより学びになるだろう)。最近の解析ソフトは簡単に統計値を算出するし、計算方法を知る必要はないかもしれないが、定義上の特徴(長所や短所)は使い手として理解しておきたい。解析ソフトは計算をするだけであり、計算の意味を考えているわけではない。自分の見たいものを、どの指標を使って示すのがよいかは、使い手が判断しなければならない。

たとえば、二つの事象の関係を捉える指標に相関係数がある。相関係数の大きさや、その統計的有意性は、必ずしも因果関係を示さない。相関関係を単に数字で捉えると忘れがちだが、具体的な事象をあてはめると気がつく。たとえば、警察官の数と犯罪の数には強い正の相関が観察されるが、だからと言って、一般には警察官を増やすと犯罪が増えるとは考えにくい。犯罪が増えればそれを抑制するために警察官が増やされる。そうであれば、因果関係は逆だ。

因果関係の把握は、政策を考える時に重要となる。先の例で言うならば、犯罪数と警官数に正の相関があるとしても、犯罪を減らすために警察官を減らすとは考えないだろう。また、たとえば、ある政策と貧困率の低さに強い負の相関があるとしても、これが因果関係でない場合には、この政策を採用しても貧困削減にはつながらない。

5. 統計データは社会科の学びを楽しくするもの

もう一度まとめておこう。人や社会を統計データによって表そうとすれば、統計学の知識だけではうまく行かない。データの使い手が社会背景に関する知識を持っていることが大切である。数学や統計学で学ぶ知識と、実際の社会の理解が融合した時に、統計をうまく使いこなすことができる。

そして、データを使う体験の前に、統計学の知識を習得しておく必要はない。むしろ、好奇心旺盛な若い時に、データを使って考える楽しみを体験させたい。必要な統計知識は、後で(必要な時に)学べばもっと社会科を楽しめると思う。

最後に、極めて個人的な話で終わりとしたい。私は高校の頃、社会科の授業が一番好きだった。でも得意ではなかった。覚えることが役立つと思えなかったからだ(働き始めて、様々な国や地域に行くようになり、ちゃんと覚えておけばよかったと後悔している)。得意だった数学や物理よりも社会科が好きだったのは、先生の話が面白かったからだと思う。余談だけど……で始まる先生の話から、見たことのない時代の、行ったことのない場所の、会ったこともない人のことを考えるだけでワクワクした。そんな私に教えてあげたい。統計データを使ったらもっと驚くことがある。誰も気づいていないことを発見できたりもする。統計データは、社会科学の学びを豊かにしてくれる。

【引用文献】

斎藤修(2006)「農民の時間から会社の時間へ：日本における労働と生活の歴史的変容」『社会政策学会誌』15, pp.3-18.