【シンポジウム：デモクラシーを考える】報告Ⅲ

デモクラシーと統計
――日本における統計の歴史と統計的な考え方――

貫　名　貴　洋

目　次

1. はじめに

近年の日本社会において，統計，ビッグデータ，データサイエンスといった用語が多用されている。2013 年に発刊された西内啓氏の『統計学は最強の学問である』は，統計学が一般社会の人々にも身近な存在になる 1 つの転機となった。当時，ビジネス書ランキングで常時上位を位置するベストセラーとなり，ビジネス誌に限らず一般誌などでも幾多の統計学特集が組まれるなど，空前の統計学ブームを招く引き金となった。

また，20 世紀終盤以降，コンピュータやインターネットの進化に伴い，情報収集の利便性が著しく向上し，統計情報を誰しもが比較的安価で，かつ，容易に利用できるようになった。こうした社会の変容に伴い，政治や行政の現場でもこれまで以上に統計•統計学の重要性や必要性が議論されている。例えば，エビデンス（科学的根拠）を重視した政策立案である EBPM（Evidence-Based Policy Making）が国内外問わず推進されている。EBPM は，政策が効果にどのように影響を及ぼしているか，統計に限ったものではなく，因果関係を把握し

たうえで講じることが必要とされている(1)。

本稿では，令和という新時代，2020年代の幕開けを迎えた今，デモクラシーと統計の関係性を見直したうえで，私たちの暮らしに必要不可欠な統計的な考え方や手法を示唆していきたい。

2. デモクラシーと統計

1868年の明治維新を迎えるまで，日本には「統計」という用語は存在していなかったとみられている(2)。「統計」とは，幕末から明治維新期に西欧から移入された，“Statistics”の訳語である(3)。現代に生きる私たちからすれば，“Statistics”を「統計」という用語として当然のように使うことができているのは，先人たちの功績によるものが大きい(4)。官庁の文書に「統計」が最初に用いられたのは，1870年（明治3年）8月4日に外務省が諸省に廻達した文書（外国貿易品輸出入の物品高表を編集，貿易年表を出版する旨を通知した文書）とみられ，その中で「統計年鑑」という用語が出ている(5)。すなわち私たちが現在過ごしている2020年は，「統計」という用語が初めて公に用いられてから，ちょうど150年を迎える歴史的な年なのである。

またもう一つ，2020年は統計史上において歴史的な年であるというトピックスにも触れておく。わが国の統計の根幹を担う「国勢調査」(6)がある。国勢調査とは，調査年の10月1日時点で日本に住んでいるすべての人と世帯を対象とする国の最も重要な統計調査であり，国勢調査から得られる様々な統計は，国や地方公共団体の政治・行政において利用されることはもとより，民間企業や研究機関でも広く利用され，そのような利用を通じて国民生活に役立てられている(7)。この国勢調査の第1回が実施されたのが1920年であり，本年でちょうど100年目を迎えることとなった。まさに日本の「統計」の歴史は，日本の「民主主義」の歴史とともにしているのである。

一方，近年において日本の「統計」は大きな危機を迎えている。先に述べた国勢調査は，統計法に基づいた国の重要な統計調査である「基幹統計調査」と

して実施され，「我が国に住んでいるすべての人と世帯を対象」としている。しかしながら，2000 年前後より個人情報保護への国民の意識が高まり，調査員が世帯に会えない状況や，調査員が世帯に会えても協力が得られないといった事例が多く発生し，調査が難航している(8)。国勢調査の結果は，衆議院議員総選挙における選挙区の改定，子育て支援や高齢者福祉対策，防災計画の策定などといった行政施策に利用されているだけでなく，企業の経営計画に用いられたり，われわれの学術研究の基盤資料として用いられたりするなど，私たちの暮らしに密接に関連づいた幅広い活用がなされている。さらに国勢調査の結果は，他の統計調査の母集団の役割をもち，標本選定の基準となっている。国勢調査の難航は，標本調査である「家計調査」や「小売物価統計調査」にも大きな影響を与えることは想像にたやすく，上記 2 調査から算出される「消費者物価指数」が実態と乖離することも考えられる。すなわち，国勢調査による結果が不正確であるという状況が恒常化してしまうと，無意識のうちに私たちの生活に不利益を生じてしまう恐れがある。

また，厚生労働省の「毎月勤労統計調査」が不適切な処理によって公表されていたことが 2018 年末に発覚し，政府が作成・公表する統計資料の根幹が揺るぐ事態となった。毎月勤労統計調査は，本来ならば従業員 500 人以上の事業所では全数調査しなければならない。しかしながら，全数調査でなくても適切な復元処理（データ補正）がされる限り，統計の精度が確保できると考えた厚労省の職員によって，2004 年 1 月調査分から東京都での大規模事業所を抽出調査に切り替えた。さらに悪いことに，この復元処理がなされていなかったというのである。復元処理がなされていなかったことより，平均賃金が低く公表されていた。こうした不正処理を意図的ないし組織的に隠ぺいしようとしていたのかどうかを追及することも重要な課題ではある。それ以上に問題視すべきは，統計法に基づく調査を，数名の公務員による自己都合的な発案，さらに不適切な処理と理解できず継続していたことに対する責任である。日本の公的統計の正確性や信憑性に泥を塗る行為であり，公的統計の意義や重要性に対する意識の低さを露呈したものであり甚だしい怠慢である(9)。

デモクラシーをどのように論ずるかという命題に対して本稿では明確な議論を避けるが，デモクラシーの根底には私たち一般国民も統計の重要性を改めて認識する必要がある。また，統計的な考え方を熟知した民意を基に，政治・行政に対して現状の課題を認識させその修正への対応を迫る必要性もあろう。現代社会を生き抜くリテラシー能力を「読み・書き・そろばん」と定義するならば，この 3 点に「統計」を加えてもよいのではないか。令和におけるデモクラシーが私たちの身近な存在であり生きたものとするためにも，「統計」の重要性を各個人が認識をした上で，誰しもが統計的な考え方を身につけることによって，改めて私たちの手中にデモクラシーが収まっていくといえよう。

3. 統計的な考え方と教育

これまで述べてきた「統計」とは，社会の実状を表す情報であり，その情報を収集した結果であり，それらを効率的に記述した「データ」そのものである。こうした「統計」をもとに，情報の分析を行ったり，意思決定において利用されたりする。その際重要となる考え方が「統計学」となる。20 世紀終盤以降，コンピュータ技術の発展により，誰しもが「統計」を用いて計算を行ったり分析をしたりという行為が容易に実践できるようになった。

こうした時世の変化を受け，『平成 20, 21 年改訂 学習指導要領』より，統計教育が明確に学習内容に組み込まれた。中学 1 年数学では，「目的に応じて資料を収集して整理し，その資料の傾向を読み取る能力を培う」という目標が定められ，平均値・中央値・最頻値・相対度数・範囲・階級といった統計用語を習得する。中学 3 年数学では，「母集団から標本を取り出し，その傾向を調べることで，母集団の傾向を読み取る能力を培う」という目標が定められ，全数調査や標本調査の意味を学習する。それまでの学習指導要領と比較すると，明確に「統計」に力点を置かれたのである。現在の大学生であれば，こうした統計教育を一通り学んでいることとなる。20 歳代後半より上の世代と，統計的な考え方を習熟しているかどうかで差がつく可能性もある。

4. 私たちの暮らしと統計

4. 1. 選挙の開票速報

政治と統計を関連付けるものとして，最も頭に浮かびやすいのは，国政選挙が実施されるたびにテレビ等で報道される「開票速報」ではないだろうか。現在では，選挙期間中の取材や投票所での「出口調査」といった事前の情勢分析に力を入れることにより，投票締め切り時点の開票率 0% の状況にあっても，大勢の「当確予想」が出されるということは周知のとおりである[10]。また，過去の開票速報[11]や，選挙結果が拮抗すると予測されている場合など，開票の途中経過を十分に分析しながら「当選確実」が出されている。こうした，開票の途中経過を踏まえながら「当選確実」を出す・出さないの決断をするために必要な知識が，統計学における「比率の区間推定」となる。信頼率 95% において「比率の区間推定」を行うための式は次の通りとなる。

$$p_i \pm 1.96 \times \sqrt{\frac{p_i \times (1-p_i)}{n}} \quad \cdots\cdots ①$$

なお，①式における p_i は各候補者の得票率，n は全投票数とする。今回は信頼率 95% としたため，計算式の中に 1.96 という数字が用いられている[12]。

4. 1. 1. 架空選挙における開票速報

開票速報において「当選確実」を出す一例を，架空選挙を用いて示したい。ここでの架空選挙は，立候補者の中で最も多くの得票を得た候補者が当選となる，小選挙区での選挙を想定している。ある選挙区には，5,000 人の有権者が存在する。その選挙区では，A 氏，B 氏，C 氏の 3 氏が立候補した。期日前投票ならびに投票当日合わせて有権者の 60% が投票をし，全ての投票が有効票とみなされた。すなわち，有権者 5,000 人に対して 60% の投票率であるので 3,000 人が投票をしたこととなる。全投票が有効票となるため，有効投票数は 3,000 票となる。開票所では一部の有権者やマスコミが立ち合う中，即日開票

が実施され，随時開票状況が公表される。その公表経過が**表 1** となる。

表 1　架空の選挙による開票結果（得票数）（有権者数 5,000 人，投票率 60%）

立候補者	開票率			
	30%	60%	90%	100%
A 氏	400	800	1,200	1,330
B 氏	350	700	1,050	1,170
C 氏	150	300	450	500
開票数	900	1,800	2,700	3,000

一般的に開票所では，表 1 のような開票率とその時点における得票数のみが公表されることが多い。しかしながら，先に記述した「比率の区間推定」では，「得票率」が必要となる。「得票率」とは。各候補者の得票数が全投票数に占める割合となる。表 1 をもとに，開票率ごとにおける各候補者の得票率を算出した（**表 2** 参照）。次項以降では，表 2 の数値を①式にあてはめて当選確率をシミュレーションしていく。

表 2　架空の選挙による開票結果（得票率）（有権者数 5,000 人，投票率 60%）

立候補者	開票率			
	30%	60%	90%	100%
A 氏	44.44%	44.44%	44.44%	44.33%
B 氏	38.89%	38.89%	38.89%	39.00%
C 氏	16.67%	16.67%	16.67%	16.67%
開票数	100.00%	100.00%	100.00%	100.00%

4. 1. 2.　開票率 30% 時点における推定

まず A 氏の得票率を推定する。開票率 30% 時点における A 氏の得票率は 44.44% であった。①式の p_i に A 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$41.20\% \leq p_A \leq 47.69\%$$

同様に B 氏の得票率を推定する。開票率 30% 時点における B 氏の得票率は 38.89% であった。①式の p_i に B 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$35.70\% \leq p_B \leq 42.07\%$$

引き続き C 氏の得票率を推定する。開票率 30% 時点における C 氏の得票率は 16.67% であった。①式の p_i に C 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$14.23\% \leq p_C \leq 19.10\%$$

このように得られた推定値をもとに，各氏の予想最終得票数を投票総数から算出すると以下の通りとなる。

$$1{,}235.94 \leq A \leq 1{,}430.73$$

$$1{,}071.12 \leq B \leq 1{,}262.22$$

$$428.96 \leq C \leq 573.04$$

この数字から当確予想を検討してみる。A 氏の予想最終得票数の下限値は 1,235.94 票，上限値は 1,430.73 票となる。B 氏の予想最終得票数の下限値は 1,071.12 票，上限値は 1,262.22 票となる。C 氏の予想最終得票数の下限値は 428.96 票，上限値は 573.04 票となる。A 氏と B 氏の予想最終得票数を見比べてみると，B 氏の予想上限値が A 氏の予想下限値を上回っているため，逆転の可能性が残されている。よって，この時点での当選確実は出せる状況にない。なお，C 氏の予想上限値は，A 氏・B 氏の予想下限値よりも大幅に下回っているため，逆転の可能性は極めて低い数字となっている。

4. 1. 3.　開票率 60% 時点における推定

前項と同様に，まず A 氏の得票率を推定する。開票率 60% 時点における A 氏の得票率は 44.44% であった。①式の p_i に A 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$42.15\% \leq p_A \leq 46.74\%$$

同様に B 氏の得票率を推定する。開票率 30% 時点における B 氏の得票率は

38.89% であった。①式の p_i に B 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$36.64\% \leq p_B \leq 41.14\%$$

C 氏に関しては，開票率 30% 時点で逆転の可能性が低くなったことにより，ここでは推定の計算を省略する。

このように得られた推定値をもとに，A 氏・B 氏の予想最終得票数を投票総数から算出すると以下の通りとなる。

$$1{,}264.47 \leq A \leq 1{,}402.20$$

$$1{,}099.10 \leq B \leq 1{,}234.23$$

A 氏の予想最終得票数の下限値は 1,264.47 票，上限値は 1,402.20 票となる。B 氏の予想最終得票数の下限値は 1,099.10 票，上限値は 1,234.23 票となる。開票率 30% 時点での予想最終得票数と見比べてみると，推定のもととなる得票率は A 氏 44.44%，B 氏 38.89% と全く同数であったのに，区間推定によって得られた下限値・上限値の数値幅が狭くなったことに気づくだろう。さらに A 氏と B 氏の予想最終得票数を見比べてみると，B 氏の予想上限値が A 氏の予想下限値を下回っているため，逆転の可能性が失われてしまった。よって，A 氏への当選確実が出せる状況が生まれる。これ以降，B 氏の得票率が急激に上昇するかしないかを見極めながら，第一報の決断が迫られる[13]。

4. 1. 4. 開票率 90% 時点における推定

同様に，まず A 氏の得票率を推定する。開票率 90% 時点における A 氏の得票率は 44.44% であった。①式の p_i に A 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$42.57\% \leq p_A \leq 46.32\%$$

同様に B 氏の得票率を推定する。開票率 30% 時点における B 氏の得票率は 38.89% であった。①式の p_i に B 氏の得票率を代入し計算すると，以下の推定結果が得られる。

$$37.05\% \leq p_B \leq 40.73\%$$

このように得られた推定値をもとに，A 氏・B 氏の予想最終得票数を投票総数から算出すると以下の通りとなる。

$$1{,}277.10 \leq A \leq 1{,}389.56$$

$$1{,}111.50 \leq B \leq 1{,}221.83$$

A 氏の予想最終得票数の下限値は 1,277.10 票，上限値は 1,389.56 票となる。B 氏の予想最終得票数の下限値は 1,111.50 票，上限値は 1,221.83 票となる。A 氏と B 氏の予想最終得票数を見比べてみると，開票率 60% 時点と同じく，B 氏の予想上限値が A 氏の予想下限値を下回っている[(14)]。

開票率 60% 時点での予想最終得票数と見比べてみると，さらに，区間推定によって得られた下限値・上限値の数値幅が狭くなった。つまり，開票率が上昇すればするほど，予想最終得票数の精度が上がるのである。またこの時点における，A 氏の得票は 1,200 票，B 氏の得票は 1,150 票と，150 票差がついていることからであり，残り 300 票のうち 50% 以上の得票を B 氏が獲得しない限り逆転は起こらない。さらに，残票を仮に B 氏の票にすべて振り替えたとしても A 氏の得票数を超えられないことが明確になった瞬間，A 氏は当選確実から当選へと表記が変わることとなる。

4. 2. 新型コロナウイルスの陽性反応および新規陽性者数の推移

本稿での議論は医学的な見解ではなく，統計学的な考え方によって得られる一考察にすぎないことをご了承の上読み進めていただきたい。

2020 年の 1 年間は，言わずもがな，「コロナ一色」となってしまった。マスコミやネットニュースは,「新型コロナウイルスの新規感染者数」を声高に日々報道し続けた。この「新規感染者数」の根拠は,「PCR 検査によって陽性反応が示された数値」であろう。2019 年末〜 2020 年初に中国にて未知のウイルスが猛威を振るっていると初めて報道されてから，ロックダウンや緊急事態宣言が発令されるに至るまで，新規陽性者数ならびに死者数が報道されることによって，日本も例にもれず世界中が恐怖の底に陥れられた。

しかしながら，緊急事態宣言が解除され，新しい生活様式下での経済活動が

取り戻されようとした以降，PCR 検査では陽性と判定されながらも症状が全く現れない「無症状」という事例も数多く認識されてきた。また，一度は陽性と判定されていたにもかかわらず，日を置かず陰性であったと再判定され，最初の判定が「偽陽性」であったとされる例もいくつか報道された。

また，PCR 検査による新規陽性者数は，週初めの月曜日には数が少なく報告され，週半ばの水曜日〜金曜日に数が多く報告される傾向にある。それにもかかわらず，報道では新規陽性者数の推移と，「前日比」を注視することによって動向を伝えているようにうかがえる。

図 1 は，東京都における新型コロナウイルスの新規陽性者数の推移である。マスコミやネットニュースの報道では，この図とほぼ同様のものが用いられている。このままでは先に述べた曜日別の特性を考慮されておらず，日ごとの陽性者数をただただグラフ化したものに過ぎない。事実，図 1 を見てわかるように，その日に報告を受けた新規陽性者数である棒グラフの高さは，日々上昇・下降を繰り返している。

以下では，新型コロナウイルスの PCR 検査による陽性反応について，また，新規陽性者数の推移について，統計学的な見地から考察を深めていく。

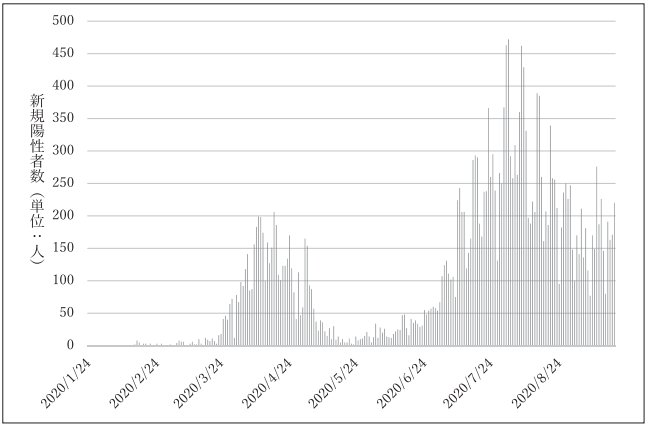

図 1　東京都における新型コロナウイルス新規陽性者数の推移[15]

4. 2. 1. 新型コロナウイルスの陽性反応について

新型コロナウイルスに関して，陽性であるか否かを検査する代表的なものが PCR 検査であろう。公表されている PCR 検査の特徴に関して簡単にまとめると次の通りとなる(16)。

（ア） PCR 検査の「感度」は 70% 程度と言われている

（イ） PCR 検査の「特異度」は 1% 程度と言われている

なお，感度とは「実際に疾患があるときに，正しく陽性が出る確率」のことを言い，特異度とは「実際に疾患がないにもかかわらず，誤って陽性が出る確率」のことを言う。間違って理解してほしくないのは，症状の有無にかかわらず行った PCR 検査での結果で陽性が示されたからといって，即陽性と判断できる材料にはならないということである。

また，日本における新型コロナウイルスの陽性者累計は約 100,000 人であり，日本人がウイルスにかかっている確率（罹患率）は日本人総人口である約 120,000,000 人のうち 0.1% 未満となる。この数字は世界的にも極めて低い(17)。次に計算する事例のために, 日本人がウイルスにかかっている確率(罹患率)を，あえて実際よりも高めの数値である 1% としておく。

（ウ） ウイルスにかかっている確率（罹患率）は 1% とする

以上の（ア）•（イ）である PCR 検査の特徴, および（ウ）の罹患率を用いて，症状の有無にかかわらず行った新型コロナウイルスの PCR 検査にて陽性反応が出た場合の，真にウイルスにかかっている確率を求める。この際必要となる公式が「ベイズの定理」である(18)。

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A)+P\left(B\middle|\bar{A}\right)\cdot P(\bar{A})} \cdots\cdots②$$

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)} \cdots\cdots③$$

②式の分子から説明する。$P(B|A)$ とは,「ウイルスを持っている状態（A）で，かつ，陽性反応が出る（B）確率」である。これに「罹患率」である $P(A)$ を掛け合わせることによって，$P(B|A)\cdot P(A)$ の値，すなわち「ウイルスを持っ

ていて，なおかつ陽性である確率」が求められる。

次に分母を説明する。$P(B|A)\cdot P(A)$ は前述のとおりである。$P(B|\bar{A})$ とは，「ウイルスを持っていない状態（$\bar{A}$）にもかかわらず，陽性反応が出る（B）確率」である。これに「ウイルスにかかっていない確率（$1-P(A)$）」である $P(\bar{A})$ を掛け合わせることによって，$P(B|\bar{A})\cdot P(\bar{A})$ の値，すなわち「ウイルスにかかっておらず，なおかつ陽性である確率」が求められる。分母である $P(B|A)\cdot P(A)+P(B|\bar{A})\cdot P(\bar{A})$ とは，「ウイルスを持っていて，なおかつ陽性である確率」と「ウイルスにかかっておらず，なおかつ陽性である確率」であり，「陽性反応を示したすべての確率」$P(B)$ となる。

よって「ベイズの定理」で求められる最終的な値 $P(A|B)$ とは，「ウイルスを持っていて，なおかつ陽性である確率」$P(B|A)\cdot P(A)$ を「陽性反応を示したすべての確率」$P(B)$ で割った値となる（③式参照）。

それでは，（ア）～（ウ）の数値を②式に代入し計算を行う。

$$P(B|A)\cdot P(A)=70\%\times 1\%=0.007$$

$$P(B|\bar{A})\cdot P(\bar{A})=1\%\times 99\%=0.0099$$

$$P(B|A)\cdot P(A)+P(B|\bar{A})\cdot P(\bar{A})=0.007+0.0099=0.0169$$

$$\frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A)+P(B|\bar{A})\cdot P(\bar{A})}=\frac{0.007}{0.007+0.0099}=\frac{0.007}{0.0169}=41.42\%$$

以上の計算結果より，症状の有無にかかわらず行った PCR 検査での結果で陽性が示された場合，真に新型コロナウイルスに感染している可能性は 41.42% となる。PCR 検査によって陽性と結果が出た場合，100% 感染しているとか，当初示した「感度」の 70% 感染しているとかと思っていたよりもかなり低い数値となっていることに注目されたい。

さらに「ベイズの定理」の便利な方法を示しておきたい。先の計算では（ウ）に該当する「罹患率」$P(A)$ を 1% として用いた。「ベイズの定理」ではこの値のことを「事前確率」と呼ぶ。そして計算によって得られた確率を「事後確率」と呼ぶ。2 度続けて同じ検査を実施する場合，「1 度目の事後確率」を「2 度目

の事前確率」に置き換えることができる。

それでは，1 度目の PCR 検査によって陽性反応が示された後，引き続き同じ検査をすることとし，連続して陽性反応が示されたとする。「ベイズの定理」に基づき，この場合の計算をしてみよう。まず，罹患率$P(A)$が「2 度目の事前確率」として，先ほどの計算結果である 41.42% に置き換わる。よって，$1-P(A)$である$P(\bar{A})$も 58.58% と置き換わる。

$$P(A) = 41.42\%$$

$$P(\bar{A}) = 1-P(A) = 58.58\%$$

（ア）•（イ）の数値, 並びに（ウ）から「2 度目の事前確率」として置き換わった数値を，改めて②式に代入し計算してみる。

$$P(B|A)\cdot P(A) = 70\% \times 41.42\% = 0.2899$$

$$P(B|\bar{A})\cdot P(\bar{A}) = 1\% \times 58.58\% = 0.0059$$

$$P(B|A)\cdot P(A) + P(B|\bar{A})\cdot P(\bar{A}) = 0.2899 + 0.0059 = 0.2958$$

$$\frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A) + P(B|\bar{A})\cdot P(\bar{A})} = \frac{0.2899}{0.2899 + 0.0059} = \frac{0.2899}{0.2958} = 98.02\%$$

最終的に得られた数値を見て，大きく様子が変わったことがお分かりいただけるだろう。症状の有無にかかわらず 2 度続けて PCR 検査で陽性反応が示された場合，真に新型コロナウイルスに感染している可能性は 98.02% となるのである[19]。

このような根拠から，1 度目の「陽性反応」のみで「感染者」と断定することには無理な事例が多く含まれているのではないだろうか。「濃厚接触者」や「無症状者」に対して PCR 検査を実施し 1 度の PCR 検査で陽性反応が出たとしても, それは「陽性反応」が出たにすぎず,「感染者」と断定するよりも「偽陽性」を疑う必要もある。「無症状」であっても，2 度続けて陽性反応が出た場合には大いに「感染」を疑うべきである。さらに深刻なケースは，事前に何かしらの症状を発症している者が，1 度目の PCR 検査で「偽陰性」（本来ウイルスを持っているにもかかわらず，検査で誤って陰性と反応が出る）反応が示される

ケースである。「感度」が 70% である限り,「偽陰性」は 1 度目の検査では 0.3% の確率で発生してしまう。こうしたケースはどのように考慮されているのだろうか。実は本稿では，これまで述べてきたことを理由に，一般的に「感染者」と称されているものをあえて「陽性者」と記述している。説明が前後してしまったが，その点をご理解いただければ幸いである。

4.2.2. 新型コロナウイルス新規陽性者数の推移について

先にも述べたが，日々報道されている新型コロナウイルス新規陽性者数の推移について，報告される曜日の特性が少なからず見受けられる。週初めの月曜日には数が少なく報告され，週半ばの水曜日〜金曜日に数が多く報告される傾向があると考えるのは自然だろう。このような特性があるにもかかわらず，日々報告される数字の大小に一喜一憂したり，前日比の数字を大々的に取り扱ったり，ましてや○日間連続とか○日ぶりなどという表現を用いて何がわかるといえるのだろうか。

こうした曜日の特性を簡単に平準化する手法として，移動平均を用いることが多い。移動平均とは，時系列データを扱う際に幅広く利用されている。一般的には，時系列データの時点のデータとして，その時点と隣接する前後数期間のデータの平均値をとったものである。一般的な 3 項移動平均であれば，

$$\bar{x}_t = \frac{x_{t-1} + x_t + x_{t+1}}{3}$$

の数式で表せる。また，一般的な 7 項移動平均であれば，

$$\bar{x}_t = \frac{x_{t-3} + x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2} + x_{t+3}}{7}$$

の数式で表せる。しかし，ここでは，直近 7 日間の移動平均を求める必要があり，木曜日の移動平均をどのように計算するかを数式化してみる。

木曜日の移動平均

$$= \frac{\text{前週金曜日のデータ} + \text{前週土曜日のデータ} + \cdots + \text{当週水曜日のデータ} + \text{当週木曜日のデータ}}{7}$$

先述 2 式と異なり，木曜日の前後 7 日間を平均するのではなく，木曜日を基点に過去 7 日間さかのぼったデータの平均を求めようとしている。この考え方に従い，直近 7 日間の移動平均を④式と定義づける。

$$\bar{x}_t = \frac{x_{t-6}+x_{t-5}+x_{t-4}+x_{t-3}+x_{t-2}+x_{t-1}+x_t}{7} \quad \cdots\cdots④$$

東京都内で初めて陽性者が報告された 2020 年 1 月 24 日から 2020 年 9 月 18 日までの東京都における新型コロナウイルスの新規陽性者数の推移を，④式に基づいた直近 7 日間移動平均値として算出し，グラフ化したものが**図 2** となる。

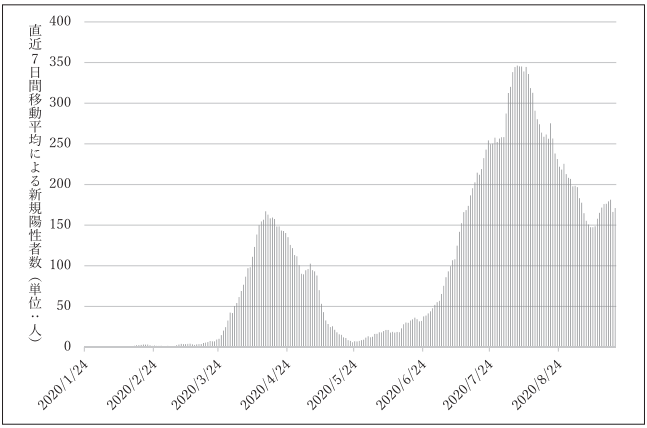

図 2　東京都における新型コロナウイルス新規感染者数の推移（直近 7 日間移動平均）(20)

図 1 と図 2 を見比べていただこう。図 2 は図 1 に比べて，新規陽性者数である棒グラフの高さが平準化され，陽性者数の上昇・下降といった傾向がわかりやすく表示されている。④式から算出される数値を利用して，前日比や〇日連続といった表現を用いることについては，原データを用いるよりもわかりやすく説明ができる。改めて④式を見ていただければわかるように，直近 7 日間のデータを平均するだけの非常にシンプルな計算方法である。この手法によって得られる数値をもとに局面の分析に役立ててほしいと願うのはもちろんである

が，簡易な計算によって私たち一個人にも理路整然と問題の焦点を照らしているのである(21)。

5. おわりに

本稿ではデモクラシーと統計について，歴史的な側面ならびに統計的な考え方の重要性について詳細に論じてきた。わが国におけるデモクラシーと統計の歴史は，明治維新期以降ほぼ同年の歩みをしている。デモクラシーについて，これまでにも多くの議論がなされ，誰しもがその重要性について認識していることであろう。ところが統計となると，自分事のように考えることが難しく，どこか雲の上の存在であるかのように勘違いしてしまうこともあるだろう。統計を知ること，統計学的な考え方を身につけることは，もっと私たち１人１人がその重要性に気づき，日常生活においても自分事として捉える必要があることにつながる。

また，本稿ではわかりやすい事例として，いくつかの統計学的な考え方を紹介した。デモクラシーと統計という，一見すれば何のつながりがないように感じられた２つの考え方が，いかに密接に関連付けられ，重要に結びついているかが１人でも多くの方に理解していただければと切に願っている。もちろん統計学は，もっと幅の広い，もっと奥の深い学問である。あらゆる人々が統計学の学びに興味を持っていただき，デモクラシーのさらなる発展に一助を担うことになるのであれば，本稿が何かしらの意味を持ってくることになるのであろう。

〈引用・参考文献〉

（図書・論文資料）

国友直人，山本拓編（2019），『統計と日本社会』，東京大学出版会。
西内　啓（2013），『統計学が最強の学問である』，ダイヤモンド社。
大橋　弘（2020），『EBPM の経済学　エビデンスを重視した政策立案』，東京大学出版会。

宮沢公男（2017），『統計学の日本史　治国経世への願い』，東京大学出版会。
溝口敏行（2003），『日本の統計調査の進化― 20 世紀における調査の変貌―』，渓水社。
涌井貞美（2013），『図解・ベイズ統計「超」入門』，SB Creative。

（インターネット資料）

厚生労働省（2019a），「毎月勤労統計調査を巡る不適切な取扱いに係る事実関係とその評価等に関する報告書について」，https://www.mhlw.go.jp/stf/newpage_03321.html，最終閲覧日：2020 年 9 月 18 日。
厚生労働省（2019b），「毎月勤労統計調査を巡る不適切な取扱いに係る事実関係とその評価等に関する追加報告書について」，https://www.mhlw.go.jp/stf/newpage_03758.html，最終閲覧日：2020 年 9 月 18 日。
日本アンチ・ドーピング機構，https://www.playtruejapan.org/code/rule/testing.html，最終閲覧日：2020 年 10 月 31 日。
日本疫学会，「新型コロナウイルス関連情報特設サイト」，https://jeaweb.jp/covid/qa/index.html，最終閲覧日：2020 年 10 月 31 日。
奥積雅彦（2018a），「公文書で初めて統計の用語が登場したのはいつか？」，『統計図書館ミニトピックス No.1』，https://www.stat.go.jp/library/pdf/minitopics1.pdf，最終閲覧日：2020 年 9 月 12 日。
奥積雅彦（2018b），「福沢諭吉，学問のすすめで統計の重要性を主張」，『統計図書館ミニトピックス No.4』，https://www.stat.go.jp/library/pdf/minitopics4.pdf，最終閲覧日：2020 年 9 月 12 日。
奥積雅彦（2018c），「なぜ「Statistics」は「統計」なのか？―「統計」の訳字が定着するまでの経緯と森鴎外」，『統計 Today　No.136』，https://www.stat.go.jp/info/today/136.html，最終閲覧日：2020 年 9 月 12 日。
SankeiBiz（2020），「国勢調査の未回収率が増加　プライバシー意識が向上，コロナの影響必至」，https://www.sankeibiz.jp/macro/news/200901/mca2009010500004-n1.htm，最終閲覧日：2020 年 9 月 12 日。
総務省，「平成 29 年 10 月 22 日執行　衆議院議員総選挙・最高裁判所裁判官国民審査速報資料」，最終閲覧日：2020 年 9 月 12 日。
総務省統計局，「令和 2 年国勢調査」，https://www.stat.go.jp/data/kokusei/2020/index.html，最終閲覧日：2020 年 9 月 12 日。
東京大学保健センター，http://www.hc.u-tokyo.ac.jp/covid-19/tests/，最終閲覧日：2020 年 10 月 31 日。

東京都新型コロナウイルス感染症対策サイト，https://stopcovid19.metro.tokyo.lg.jp/，最終閲覧日：2020 年 9 月 18 日。
World Health Organization，“WHO Coronavirus Disease（COVID-19）Dashboard”，https://covid19.who.int/，最終閲覧日：2020 年 10 月 31 日。

注

（1） 大橋（2020），P.1 参照。
（2） 奥積（2018a）参照。
（3） “Statistics” がまだ「統計」という一般的な訳語となっていなかった時代の使用例を挙げる。1872 年（明治 5 年）2 月初出である，福沢諭吉の『学問のすすめ』第 13 編に，「その数を記したるスタチスチクの表ありて」とある。なお，福沢諭吉は，この前後のくだりにおいて，統計的なものの見方の重要性を説いている（奥積（2018b）参照）。
（4） “Statistics” が「統計」という訳字が定着するまでの論争については，奥積（2018c）によって詳細にまとめられている。
（5） 奥積（2018a）参照。
（6） 国勢調査は，英語では Census や Population Census と呼ばれる。Population とは一般的に人口と訳されることが多いが，母集団という意味もある。つまり国勢調査とは，「人口調査」でもあるが，「母集団調査」という側面も担っている。なお，日本の国勢調査は 5 年おきに実施されているが，世界では 10 年おきに実施されるケースも少なくない。
（7） 総務省統計局，「令和 2 年国勢調査」参照。
（8） 国勢調査の未回収率は，2000 年調査は 1.7% だったが，2005 年は 4.4%，2010 年は 8.8%，2015 年には 13.1% と年々上昇している（SankeiBiz（2020）参照）。
（9） 厚生労働省（2019a），厚生労働省（2019b）参照。
（10） 本稿では「出口調査」の統計学的見解を述べることはしないが，脚注で軽く説明をしておきたい。出口調査とは，報道関係機関が全国の投票所に調査員を派遣し，投票所出口にて，投票済みの有権者から投票した候補者や政党名，時には選挙の争点となっている意見を，対面もしくは書面にて調査する方法である。2017 年（平成 29 年）10 月 22 日に実施された衆議院議員総選挙における全国の投票所総数は 47,741 ヶ所であった（総務省，「平成 29 年 10 月 22 日執行　衆議院議員総選挙・最高裁判所裁判官国民審査　速報資料」参照）。おそらくすべての投票所，全ての投票時間帯を調査することは不可能であろうから，調査

主体が無作為抽出法（random sampling）等によって投票所や時間帯を選定し，出口調査を実施していると想像される。また，選定された出口調査の対象地点で，一定数以上の回答（標本）を得ることも求められる。もし選出方法に何らかの意図があったり標本数が一定数を満たなかったりという状況が生まれれば，出口調査から得られる予測と実際の投票結果に大きなずれが生じることとなる。

（11） 私の小学生時代，衆議院議員総選挙の「開票速報」の報道によって，「当選確実」が出されながらも最終的には落選になるという候補者がテレビ画面に映し出されたことがある。当時の衆院選は中選挙区制で，複数の有力候補者が当落接戦の状況にあり，最後の最後で都市部の大票田が開いて逆転されたように記憶している。小学生ながら，こうした間違いは絶対にあってはいけないことだという意識は既に持っており，当時から「当選確実」のメカニズムを知りたいと思うようになったのは言うまでもないことだろう。

（12） この 1.96 という数字の意味に関しては統計学の重要な考え方を理解する必要があり，大学レベルの統計学の教科書を一読することをおすすめする。

（13） 信頼率を 99% としてこの時点での 2 名の予想最終得票数を推定し直したところ，$1{,}242.68 \leq A \leq 1{,}423.99$，$1{,}077.73 \leq B \leq 1{,}255.60$ と B 氏の上限値が A 氏の下限値を上回っている。開票率 60% における A 氏の当選確実はまだ困難な状況といえる。

（14） 信頼率を 99% としてこの時点での 2 名の予想最終得票数を推定し直したところ，$1{,}259.32 \leq A \leq 1{,}407.35$，$1{,}094.05 \leq B \leq 1{,}239.28$ と，B 氏の上限値が A 氏の下限値を下回る。よって，信頼率 95% の推定結果と同様に，A 氏に当選確実を出せる状況にある。

（15） 東京都における公表資料（https://stopcovid19.metro.tokyo.lg.jp/）をもとに筆者が作成した。

（16） 東京大学保健センターや日本疫学会などの数値を参考としている。

（17） 世界各国における新型コロナウイルスの累計感染者数については，世界保健機関（World Health Organization）の公表資料（https://covid19.who.int/）にて詳細に報告されている。

（18） 「ベイズの定理」については，大学レベルの統計学の教科書であればほぼ説明がなされている。「ベイズの定理」を応用的に簡潔にわかりやすく説明しているものとして，涌井（2013）P.93-P.103 などを参照されたい。

（19） このような「ベイズの定理」の計算は，インフルエンザの簡易検査，薬局で販

売されている妊娠検査キット，スポーツ界で実施されているドーピング検査などにも同様に適用できる。インフルエンザの簡易検査や妊娠検査キットであれば，何かしらの症状があって検査に臨むことが多く，先ほどの「症状の有無にかかわらず」という条件と合致しない。よって，1度目の検査で陽性と出た場合でも十分な結果が得られている。ドーピング検査の場合には，検査対象選手がドーピング薬物を使用しているか使用していないかわからない状態での検査となるため，A検体で陽性と出た場合には引き続きB検体の検査が実施されることとなっており，上述の「ベイズの定理」の考え方が応用されていると考えられる。ドーピング検査の方法については，日本アンチ・ドーピング機構（https://www.playtruejapan.org/code/rule/testing.html）などを参照されたい。

（20） 東京都における公表資料（https://stopcovid19.metro.tokyo.lg.jp/）をもとに筆者が作成した。

（21） こうした移動平均法による時系列データの平準化は特殊な方法ではなく，日々公表される株価の推移や四半期ごとに公表される国内総生産（GDP）などにも「季節調整」として応用的に利用されている。