

「統計」あれこれ

— ナイチンゲール・森鷗外・沖縄県民投票 —

筑波総研株式会社 顧問

筑波大学名誉教授 高 木 英 明

1. はじめに

コンピュータと通信技術の劇的な進歩により、以前には考えられなかった多種多様の大量データ（有用・無用を問わず）が収集・蓄積・処理できるようになって、「ビッグデータ」が跋扈し、その中で『統計学が最強の学問である』（西内啓、ダイヤモンド社、2013年）という刺激的なタイトルの本もベストセラーになった。

しかし、「学」という字が付く「統計学」には「難しい」という印象が強く、「ビッグデータ」ほどのバズワードにはなっていなかった。

このような状況下で、昨年末に厚生労働省における「毎月勤労統計」の不正が発覚して以来、「統計」に関する話題がにわかに喧しくなってきた。

筑波銀行でも、従業員への「従業員満足度調査」と、ご来店になるお客さまへの「顧客満足度調査」をそれぞれアンケート形式で実施し、その結果の統計分析を活かした「根拠に基づく（evidence-based）」意思決定による経営が進んでいる。

本稿では、最近の不正統計問題の政治的論争には立ち入らず、統計にまつわるいくつかの話題を提供する。

2. ナイチンゲールは統計学者だった！

フローレンス・ナイチンゲール(Florence Nightingale、1820～1910)は、イギリスの裕福な上流階級の家系に生まれ、語学・芸術・科学にわたる幅広い教養を授けられたが、農民の悲惨な生活を見て、人々に奉仕する仕事に就きたいと考えるようになり、自ら志願して病院で看護の訓練を受けた。

1854年に勃発したクリミア戦争（ロシアとトルコ・フランス・イギリスとのクリミア半島を主戦場とした戦争）では、イギリスから従軍看護師団長として派遣され、野戦病院での献身的な看護活動と負傷兵の衛生状況の改善に取り組み、「クリミアの天使」と呼ばれた。これがきっかけとなり、後年、看護婦は「白衣の天使」と呼ばれ、彼女の誕生日である5月12日は「国際看護師の日」と定められている。

このように、看護師の鑑として日本でも知らない人はいないナイチンゲールは、実は、統計学においても偉大な功績を残している[文献1]。

クリミアの戦地から帰還した彼女は、統計学者のファー（William Farr）の指導を受けて野戦病院や兵士の生活環境のデータを分析し、上流階級の人脈を活かして、1858年に国会に提出したビクトリア女王の勅選委員会報告書で「戦場では負傷で死ぬ兵士よりも陸軍病院の劣悪な衛生環境の中で伝染病等により死ぬ兵士の方が圧倒的に多い」ことを明らかにした。

この報告書の付録第72には、統計結果の「見える化」が多数の図に示されているが、特に、陸軍兵士の毎月の原因別死亡率を示した「bat's wing（コウモリの翼）」と呼ばれる多色の円グラフが注目される。

彼女は1858年に「英国陸軍の衛生改革への統計的手法の活用」の功績から王立統計学会の初的女性会員に選ばれ、1874年には米国統計学会の名誉会員にも選ばれた。

3. 森鷗外と陸軍の脚気病対応

医学と統計学との関わりのエピソードは日本にもある。以下に、参考文献2に従って、森鷗外と陸軍の脚気病対応（の失敗）について述べる。

明治の文豪として知られる森鷗外（本名 森林太郎、1862～1922）は、帝国陸軍の軍医にして公衆衛生学の研究者でもあった。彼は、大学卒業後に陸軍省へ入り、1884年から約4年間ドイツに留学して衛生学や陸軍医事を学び、後に、陸軍医務局長・軍医総監にまで登りつめた。

日本では江戸時代、将軍から町民に至るまで多くの人々が脚気に悩まされ、明治時代になっても、壮年が集まる陸軍・海軍で多数の患者が発生していた。

海軍では、軍艦の乗組員に多数発生する脚気の原因について、遠洋航海に出た2隻の軍艦の乗員に異なる食事を与えて、脚気患者の発生数を比較した。その結果、食事を白米からパンなどの洋食に変えることで、脚気患者はほとんどいなくなった。

一方、陸軍では、森がドイツで行った食物の栄養研究の結果を「米は西洋食と異なることなし」と報告した。海軍との対立もあって、「脚気ウィルス説」が堅持され、白米を主食とする食事は変わらなかった。陸軍での脚気の大量発生は、日清戦争(1894～95年)・日露戦争(1904～05年)においても続き、脚気患者数は陸軍の戦病者中で最多であった。

その後、鈴木梅太郎の研究により脚気予防に有効な^{こめぬか}米糠の成分(オリザニン)が発見され(1910年)、脚気の原因は細菌ではなく、ビタミンB不足であることが判明した。

森は「海軍が行ったような統計的医学研究では因果関係は断定できず、実験的医学研究によらなければならない」と主張したが、これは、現在の「ビッグデータからは相関関係しか見つけることはできず、因果関係は究明されない」という医療関係者のデータサイエンスに対する評価に似ている。

4. 統計学は「国状学」

今回の不正統計問題に端を発する国会論戦やマスコミ報道を通して、多くの国民は、統計が国の政策決定の基盤になっている（いた）ことを認識した。しかし、統計はもともと国の状況をデータで捉える方法として発祥したものである。既に古代文明発祥の各地において、国家の経営のために行われた土地・住民・産物等に関する調査結果が記録されている。

中世に一時衰退したものの、ルネサンスを経た大航海時代には、^{いんせい}殷盛を極めた西欧の近代的国家において、国の情勢に関する諸般の事情調査が必要とされ、17世紀のドイツでStaatenkunde（国状学、国勢学）としての体系が整えられた。

現在でも、「統計学」は、英語ではstatistics、ドイツ語ではStatistik、フランス語ではstatistiqueと言い、stateは「国」「状態」等の意味である。

日本では、明治維新により西洋文明とともに移入された統計の重要性と有用性を、国家の観点から強く認識したのが、福沢諭吉（慶応義塾大学の創始者）と大隈重信（政治家、早稲田大学の創始者）である。

福沢諭吉は、『文明論之概略』の中で、統計学を文明研究の方法論として位置づけた。大隈重信は、帝国憲法と帝国議会による立憲政治の下で、国政を議論によって進めるために、統計データをその拠り所として官庁の統計制度を整備した。

大隈重信が2回目の内閣総理大臣のときに発した内閣訓令には「統計ハ国家社会各般ノ現象ヲ観察シ其消長ノ跡ヲ表顕スルモノニシテ将来施設ノ指針トスベキモノナルノミナラズ學術研究ノ基礎タルベキモノトス」と書かれている。

（一財）統計研究会 会長の宮川公男氏は、今回の不正統計問題が発覚する前に出版された著書[文献2]において、「統計の重要性に明示的に言及する大物政治家がほとんど見当たらない今日とは隔世の感がある」と嘆じている。

表1 「辺野古埋め立て」沖縄県民投票前の世論調査と県民投票の結果

	世論調査 (2月16, 17日)	県民投票 (2月24日)
賛成	180人 (16%)	114,933人 (19.10%)
反対	664人 (59%)	434,273人 (72.15%)
どちらでもない	236人 (21%)	52,682人 (8.75%)
その他・答えない	45人 (4%)	—
有効回答数	1,125人 (100%)	601,888人 (100%)

第2次世界大戦中、政府統計は情報統制の下で秘匿・歪曲された。戦後の吉田茂元首相はマッカーサー連合軍最高司令官に「戦前に我が国の統計が完備していたならば、あんな無謀な戦争はやらなかった」と伝えた後、統計機構の強化に着手して、統計委員会（のちに統計審議会）を設置し、統計法を制定した。こうして、1950年代には、日本の政府統計は世界的に高い評価を得るに至った（朝日新聞、2月18日朝刊）。

今回の官庁統計破綻の理由の1つは、度重なる行政改革の中で、職員が大幅に削減され、組織の権限も縮小したと言われるが、国民が信頼できるデータに基づく行政のために、政府が責任をもって建て直すことを求めたい。

5. 世論調査の標本サイズ

少し数値の入った話題も提供しよう。

本年2月18日（月）の朝日新聞朝刊によると、「辺野古埋め立て」の賛否に関する沖縄県民投票に先立ち、2月16, 17日に朝日新聞社が独自に行った沖縄県内の有権者を対象とした電話世論調査において、無作為に選ばれた有権者から得られた1,125件の有効回答は「辺野古埋め立て」に対する「賛成」が16%、「反対」が59%、「どちらでもない」が21%であった。

2月24日に実施された県民投票では、投票資格者数（有権者数）1,153,591人に対し、投票数605,394人（投票率52.48%）、有効投票数が

601,888人（有効投票率52.18%）となり、「賛成」が114,933人（19.10%）、「反対」が434,273人（72.15%）、「どちらでもない」が52,682人（8.75%）であった（表1）。

この結果に対する統計学的解説を試みる。

世論調査の標本サイズを見てみよう。世論調査のときの有権者も1,153,591人であったと仮定すれば、その中からたまたま選ばれた僅か1%弱に満たない1,125人の回答者による回答内訳で、県民全体の意向を判定したことになる。これではよいのだろうか？

「反対」が59%ということであるから、664人（ $1,125 \times 0.59 = 663.75$ ）が「反対」と答えたのだろうが、調査に選ばれなかった有権者（ $1,153,591 - 1,125 = 1,152,466$ ）人が全て「賛成」かもしれない。その場合には、もし全有権者に聞いた場合には、「賛成」が（ $180 + 1,152,466 = 1,152,646$ ）人であり、（ $1,152,646 \div 1,153,591 \times 100 = 99.92\%$ ）が「賛成」ということになる。

しかし、同じ論法により、もし調査に選ばれなかった有権者が全員「反対」なら、反対者は（ $664 + 1,152,466 = 1,153,130$ ）人で99.96%となる。

このような真逆の場合がともに起こり得るので、世論調査の結果である「反対59%」を単純に全対象者の意向に近いと考えるのはおかしいのではない。そもそも、新聞の読者が知りたいのは、対象者の1%に対する調査結果ではなく、全対象者の意向に関する予想であるから、新聞社は

世論調査における適切な標本サイズの算出法

母集団のサイズ	N = 調査の対象となる全員（母集団）の人数
標本のサイズ	n = 意見を聞く人たち（標本）の人数
母集団の比率	p = 母集団の人たちが「ある意見」をもっている比率
標本誤差	s = 標本の人たちが「ある意見」を示す比率のバラつき
信頼水準	r = 標本から母集団平均を推定する信頼度によって決まる定数（表2）

表2 推定の信頼度と信頼水準（ r ）の関係

信頼度	信頼水準（ r ）
90%	1.64
95%	1.96
99%	2.58

$$n = \frac{N}{\left(\frac{s}{r}\right)^2 \frac{N-1}{p(1-p)} + 1} \quad (A)$$

それを示さなければならない。

同様の疑問や不満を、内閣支持率や特定の政策に関する世論調査の結果等に対して抱いている読者もおられるだろう。確実な結果は、対象者全員に聞いてみなければ分からない。

全対象者に意見を聞く調査を**全数調査**といい、これに対して、一部の対象者を選んで聞く調査を**標本調査**という（調査に選ばれた対象者を**標本**と呼ぶ）。標本調査の結果は、何らかの合理的な理由を挙げて、全数調査の結果に近いことが示されなければならない。製造業における品質管理でも、同様の方法による製品の抜き取り検査が行われている。

全数調査の利点は正確性である。但し、意見を聞く対象は人間であるから、工業製品とは異なり、そのときの気分や聞かれた相手により、回答がぶれることも起こり得るので、翌日に同じ全数調査をすれば異なる結果が出るのが普通である。一方、欠点は、対象者の多さに伴い経費がかさむことである。従って、多くの調査では、経費削減のために、対象者の一部を選んで標本調査が行われる。

標本調査は、一部の対象者に対する調査結果から、統計学的処理により全対象者の意見を推定する手法が開発されているが、回答者が**無作為**（ランダムともいう）に選ばれることが前提とされている。例えば、ある地区の住民の平均所得を知るために、有職者だけを調査することは適切ではない。

また、世論調査では、よくコンピュータで無作為に数字を組み合わせて作成した固定電話番号に電話をかけるRDD（random digit dialing）法で回答者を選んだとしているが、ビジネスパーソンや学生が在宅していない昼間に「固定電話がある場所」にいる人たちが、性別・年齢・職業等で「平均的な国民」とは考えられないので、電話番号簿が無作為に作成されたとしても、回答者は無作為には選ばれていない。

標本調査において、標本が無作為に選ばれれば仮定するとき、調査結果から、母集団について確からしい推定を行うために、統計学における式(A)を使うことができる。

式(A)は、「各人が比率 p で『ある意見』をもつ N 人の母集団に対して、 n 人の標本を抽出してそ

の意見をもつかどうかを聞くという調査を行うと、その結果について、信頼水準 r に対応する信頼度で誤差 s を許容する判定ができる」というものである。逆に言えば、「信頼水準 r に対応する信頼度で、誤差 s を許容する判定をするためには、サイズ n の標本を使わなければならない」ということである。

式(A)から、高い信頼度(r 大)で誤差が少ない(s 小)結果を得るためには、標本サイズを大きく取る(n 大)必要があることが分かる。式(A)は簡単であるが、標本調査の目的が母集団の比率 p を推定することであるにもかかわらず、その未知な p の値を標本サイズの算定に使っているという矛盾がある。そこで、便宜的に標本調査で得られた比率を p として使うことにすると、関数 $p(1-p)$ が $p=0.5$ のときに最大となり、このとき大きめの標本サイズ n が得られるという安全側に立って、 $p=0.5$ とおくことがある。

表3に、標本サイズと標本誤差について式(A)を用いて計算した3つの例を示す。いずれの例においても、信頼度は95%とする。

まず、表1の世論調査において、「反対」が59%という結果から、県民投票の行方について予測する。表3の「例(1)世論調査」の列には、「サイズ $N=601,888$ の県民投票の有効投票を母集団とし、サイズ $n=1,125$ の有効回答を標本とする標本調査においては、95%の信頼度で、標

本比率59%に対する結果の誤差は2.87%以内である」ことが示されている。これは、県民投票の有効投票率が52.18%になれば、95%の信頼度で、有効投票者には「反対」の人が56.13~61.87%いると推定できることを示している。

しかし、実際の県民投票では有効投票数の72.15%が「反対」であったため、世論調査の結果は大きく外れたことになる。その理由は、世論調査において、「標本サイズが小さすぎた」、「標本が無作為に選ばれていなかった」等のために母集団比率とはかけ離れた標本比率の値を得てしまったからである。あるいは、公表された世論調査の結果が数日後の県民の投票行動に影響を与えたのかもしれない(朝日新聞は、自らの世論調査の信頼性について統計分析や検証を示していない)。

同様に、県民投票を標本調査と見なせば、全有権者数が母集団のサイズとなり、有効投票数が標本サイズとなる。県民投票の結果から、95%の信頼度で、全有権者の意向をどの程度の誤差で判定できるかを表3の「例(2)実際の県民投票」の列に示す。この計算結果によれば、有効投票数が601,888人の場合に、95%の信頼度で、標本誤差が0.08%であるから、全有権者について「反対」の比率は72.07~72.23%にあるという、確度の高い判定ができる。この判定が間違っている確率は5%以下である。

表3 沖縄県民投票における世論調査と有効投票のサイズ

	例 (1) 世論調査	例 (2) 実際の県民投票	例 (3) 架空の県民投票
母集団	県民投票の有効投票者	全有権者	全有権者
母集団のサイズ N	601,888 人	1,153,591 人	1,153,591 人
標本	有効回答者	有効投票者	1%以内の誤差を保証する 最少数の投票者
標本のサイズ n	1,125 人	601,888 人	7,668 人
標本の比率 p	0.59	0.7215	0.7215
信頼水準 r	1.96	1.96	1.96
標本誤差 s	0.0287	0.0008	0.0100

また、「例(3)架空の県民投票」の列には、このような標本誤差が1%ぐらい広くてもよいなら、僅か（無作為に選ばれた）7,668人の有効投票数でよいことが分かる。県民投票の日に、投票率が50%を超えたかどうかに関心が寄せられたが、全有権者の意思の推定に関しては、統計学的には無意味な懸念である。

6. サービス企業における顧客満足度の統計調査

企業における顧客満足度のアンケート調査は、通常、標本調査で行われる。筑波銀行では、顧客の声を正しく理解して業務に反映させ、地域金融機関として、より一層充実した金融サービスの提供に向けた取り組みを全行あげて行っており、そのための基礎データとして、全営業店で来店顧客（ATMのみの利用者を除く）を対象としたアンケート調査を定期的に実施している。

直近では、平成29年6月に「筑波銀行お客さまアンケート」を行った[文献3]。アンケート用紙を来店時に窓口で渡し、待ち時間を利用して回答を記入した後、回収箱に入れていただいた。

筑波銀行の全営業店の窓口には1日当たり平均約31,000人の来客があるが、今回は、全営業店に5,170枚のアンケート用紙を配付し、4,361枚を回収した。この場合の標本誤差を式(A)で計算すれば、50%の満足度を示す質問項目に対して、信頼度95%で誤差が1.38%となるので、全来店顧客の意向を推定するためには、この回収数で十分である。

我が国では、ヒトを対象とする民間のサービス企業（銀行のみならず、ホテル、レストラン、飛行機・鉄道・バス、学校、病院等）において、統計的手法による顧客満足度調査はあまり普及していないと思われる。その理由としては、以下のようことが考えられる。

- (1) 経営者にとって、顧客満足度調査は自身の経営を評価されることになるので、やりたくない。
- (2) 経営者は、顧客満足度向上よりも従業員人件

費等の経費削減に重点を置いている。

- (3) 自己の「経験と勘」に基づく経営に自信をもつ経営者は、顧客満足度が数値で測られることを信じない。
- (4) 経営者・経営企画管理部門のスタッフの多くが、「平均値」よりも難しい統計計算の数学的意味を理解できない。
- (5) 高度な統計的分析手法を提供する（高価な）ソフトウェアが統計学の言葉で出力する数字の意味を自社の経営の言葉に読み換えるのが容易ではない。
- (6) 企業の経営者・経営企画管理部門のスタッフにも理解できるような統計学やデータサイエンスの評価尺度が開発されていない。

顧客満足度調査では、数値化された満足度指標が統計的に正しい方法で示されても、それが経営者や経営企画管理部門のスタッフに理解されなければ意味がない。

さらに、それが自社の業務改善施策につながり、その結果、経営指標の向上に結び付くというサイクルを実感できるとき、統計的顧客満足度調査は、初めて企業において普及し、サービス産業全体の底上げに貢献すると思われる。

このサイクルの完結は、単純な物理法則が支配する工業製品の生産プロセスの改善よりもはるかに難しいが、今後、達成を目指すべき目標である。

参考文献

1. 丸山健夫、ナイチンゲールは統計学者だった！
ー統計の人物と歴史の物語ー、2008年6月、日科技術連出版社。
2. 宮川公男、統計学の日本史ー治国経世への願いー、2017年9月、東京大学出版会。
3. 高木英明、『筑波銀行』お客さまアンケート調査の結果報告ーお客さまの声は天の声ー、調査情報、No.47, pp.10~15, 2018年3月号、筑波総研株式会社。