

## 統計学はお嫌いですか? —「でたらめ」の使い道—

中央大学大学院 公共政策研究科教授 細野 助 博

「あいつはでたらめな奴だ」などと「でたらめ」は信頼がないと同義に使われます。でも、「でたらめ」にもちゃんとした使い道があるんです。私が昔勤務していた研究所では、限られた燃料と時間の中で領海侵犯する潜水艦を「効果的に」探索・発見するため「対潜哨戒機<sup>1)</sup>のフライト計画」のコンピュータプログラムを開発していました。コンピュータ内蔵のさいころ（擬似乱数）を使い、哨戒機が迅速かつ正確にどの経路をどの速度で飛ぶかを指定するシステムの開発です。こうしないとパイロット達は自分の好きな航路を無意識に選択するからです。

常勝将軍ほど自分の勘に頼らず、巫女やシャーマンたちが下す「神託や占い」に素直に従った故事と同じです。ですから部下は信頼して戦地に乗り込むのです。かのシーザーは、「賽は投げられた」とルビコン川を渡るときに演説します。「なんと愚かな」ではなく、まさにこれこそ先人の知恵です。

### 1. 信頼できる標本選択

#### (1) 偏りのない選択の重要性

国でもどこの自治体でも、重要な計画を立てるときに審議会や委員会を作ります。でもその委員には活動分野や各地域の代表が多く選ばれるので、多少の一般市民を入れても大体同じメンバーで構成されてしまいます。そうすると皆さんが経験しているように、議論を始める前から結果が見えてしまいます。そこで何か新鮮味をだそうと、メンバーの選び方に苦心惨憺（さんたん）することになります。

その解決策の一つが討議民主主義や熟議民主主義とか言われているもので、メンバーを「無作為＝でたらめ」に選び、正確な情報提供と十分な議論と学習機会（間違ふことの重要性、相

手の尊重）を整えて、少人数で行われる会議体です。

なぜ、でたらめを用いようとするのでしょうか。皆さん行政畑にいれば、サイレント・マジョリティと呼ばれる多様な人たちの意見を「できるだけまんべんなく」吸い上げたいですよね。では「まんべんなく」とはどんな意味でしょうか。単純にできるだけ多数の人から聞くというだけでは能がありません。「まんべんなく＝偏りなく」を実現するために知恵を働かせて、「でたらめ」を活用しようということです。「まんべんなく＝偏りなく」が保証されないと、とんでもない政策決定がなされます。

堅苦しい統計のお話より実例から紹介しましょう。米国の有名な例ですが、学歴が収入に与える影響を調べて、誰にも教育を受けさせようという計画がある州で立案されました。短絡的な担当者が貧困水準の1.5倍の収入層までを標本とし、高収入の人を標本から除外しました。その結果、教育が収入アップにつながるから教育支援を貧困層にという計画を支持する明確な分析結果が出てきませんでした。もし高収入の人も標本として抽出したら、「教育は誰にとっても収入アップにつながるから、貧困を撲滅するにはまず教育支援だ」という計画が実現したかもしれません。これを「選択バイアス」と言います。

#### (2) でたらめが保証する「偏りのなさ」

今年は米国大統領選挙の年です。さあ、オバマ氏の次の大統領は誰になるでしょう。米国大統領の決定は世界に影響大ですから、興味本位とはいきませんね。そこで昔の大統領選にまつわる話をします。

1932年に雑誌『リテラリーダイジェスト』は「次の大統領には誰?」というハガキを電話や

自動車を持つ有権者に2,000万枚（当時の人口の約1/3の人たち）送りました。集まった約300万枚の回答から、ルーズベルトの得票率57.4%をまるで「魔法みたい」に誤差0.9%でピッタリ当てました。自信をつけたその雑誌社は4年後、今度は1,000万枚のハガキを出します。そしてルーズベルト43%、ランドン57%という予測を誇らしげに発表しました。

他方、わずか3,000人の意識調査をした世論調査会社がありました。今ではすっかり世論調査で有名になったギャラップ社です。ルーズベルト54%、ランドン46%と予測しました。実際はルーズベルト60.2%、ランドン39.8%ですから、結果はギャラップ社の勝ち。データはたくさんあれば良いとは言えない教訓です。『リテラリーダイジェスト』は名声を失ってやがて廃刊されます。

対するギャラップ社の場合、枚数は少ないものの、有権者を居住地、性別、年齢別、人種別などそれぞれの構成比を注意深く観察し、3,000枚を割りふったので比較的「まんべんなく」各層の意見が得られたのです。

しかしギャラップ社も1948年は予測に失敗しました。そこで偏りのない世論を予測しようと「でたらめ＝無作為」の性質を全面的活用することに方針を転換したわけです。その結果、1980年のレーガン対カーター選では見事レーガン勝利を予測しました。統計手法を活用しようという努力と工夫がギャラップ社を救ったのです。

ところで皆さんが期待する「でたらめ」って実は奥が深いのです。例えば、でたらめに抽出するとしても、最もシンプルな「単純無作為」の抽出方法がベストの結果を出すとは限りません。サンプルをk倍に増やしても平均の推定値の正確度はkの2分の1乗しか向上しません。今はやりの熟議民主主義ではありませんが、会議の「参加者を単純無作為で選びました」と胸を張る自治体も多いのですが、努力も中途半端な結果になる例も多いようです。単純無作為で市民を1,000人選び、その結果200人参加したと

いても、条件をクリアできるのが専業主婦やリタイアした高齢者だけの偏った構成ではいただけません。

そこで、ギャラップ社が工夫したように正確度を上げるために有益な情報を取り入れた「層別無作為」抽出を採用する方がベターです。地域の違い、性別、年齢構成、職業分布などの事前の情報を活用して、構成比が高くて「散らばりが大きいグループ」のサンプル数は多く、構成比が低くて「散らばりが小さいグループ」のサンプル数は少なく抽出するという工夫です。自治体職員の皆さんも、市民に「まんべんなく」参加してもらうために、もうひと工夫をしてみませんか。

#### (3) 「でたらめさ」の見分け方

2010年、EUとIMFからギリシャの統計調査を依頼された専門家がギリシャの検察当局に起訴される事件が起こりました。彼が経済統計情報を国が関与して改ざんしたことを突き止め公表したからです。

なぜインチキが調査した専門家にばれたのでしょうか。経済のデータは一種独特です。しかし政府が発表する公式数字は、最初の桁の数字も1から9まで「均等に現れて1/10」になるわけではありません。「ベンフォードの法則」によれば、なんと1は9の6倍多く、3よりも2倍多くなるケースが多いそうです。でもギリシャ政府の公式データの数字にこの法則がまるっきりあてはまっていなかったのです。ギリシャ政府にとって「都合のよいでたらめな」数字が発表されていたのです。

今まで話してきた「無作為＝でたらめ」ではなく、ギリシャの場合は「作為的不正＝信頼のおけない＝でたらめ」ですから、2つの「でたらめ」を混同しないでください。その後のギリシャの政治的混乱や社会的混乱はニュースで見ての通りですね。そしてそれらが火種になってEUの結束までもが乱れてしまいました。まだ火種がくすぶっていることは、皆さんよくご存じの通りです。

では「無作為＝でたらめ」かどうか、どうやっ