

国勢調査・面接調査を利用した比較検討

# インターネット調査に おけるバイアス

NIRA 研究報告書

2022.3

谷口 将紀・大森 翔子 著

## Contents

<b>第1章</b>	<b>はじめに</b> .....	1
	1. イントロダクション.....	1
	2. 問題の所在.....	2
	3. 本プロジェクトの目的とリサーチクエスチョンの設定.....	3
<b>第2章</b>	<b>研究方法</b> .....	6
	1. 調査概要と記述統計.....	6
	2. 各調査データにおける質問回答分布の乖離.....	10
	3. 調査データの補正方法.....	14
<b>第3章</b>	<b>分析と結果</b> .....	19
	1. 従来型補正.....	19
	2. 国勢調査の調査票データを用いた直接補正.....	27
<b>第4章</b>	<b>結論と暫定的対策</b> .....	34
	参考文献.....	37
	付録.....	39



# 第1章 はじめに

## 要旨

インターネットの普及に合わせてインターネット調査も幅広く利用されるようになったが、インターネット調査には代表性をはじめとする問題が指摘されてきた。一方、無作為抽出に基づく従来型調査においても、回収率の低下に伴う非回答バイアスが無視できない水準に達している。ところが、国勢調査と面接調査、インターネット調査の三者を同時かつシステマティック比較した先行研究は意外に少ない。そこで本プロジェクトでは、無作為抽出に基づく面接調査とインターネット調査を同時に実施することにより、国勢調査結果を合わせて、各調査モードの回答者間にどのようなバイアスが生じているのかを観察し、調査バイアスの補正可能性を検討することにした。

## 1. イントロダクション

学術目的・それ以外の目的を問わず、インターネット調査<sup>1</sup>が盛んに利用されるようになってきている。日本マーケティング・リサーチ協会によれば、インターネット調査の市場規模は年々拡大し、各調査会社がアドホックに受託する社会調査（1回限りの社会調査）に占める割合は72%以上に達したという（日本マーケティング・リサーチ協会 2021）。社会科学分野の学術調査に限っても、インターネット調査の実施件数は増加傾向にある（三輪・石田・下瀬川 2020）。

インターネット調査は、何といてもその手軽さが売りである。調査手法によっては、数時間で質問の決定から回収——そして分析を行うことも可能だろう。迅速に、安く、簡単に人々の意識に関するデータを得ることが出来る、便利なツールであることに間違いはない。

その反面、インターネット調査は多くの課題も抱えている。例えば大隅（2002）は、①インターネットユーザーが日本人の全体を代表していない、②（インターネット調査における）登録者集団がインターネットユーザーを代表していない、③回答者が登録者集団を代表していない、と問題点を述べている。インターネット調査黎明期から指摘さ

---

<sup>1</sup> インターネット調査の他、ウェブ調査、オンライン調査など様々な呼ばれ方をしており、それぞれの意味を区別している場合もあるが、本稿では「インターネット調査」と総称し、先行研究の用語も適宜置き換えた。

れてきた、これらの——サンプルの代表性に関する——問題点は、20 年近く経過した現在においても解消されているとは言い難い。にもかかわらず、インターネット調査が広く利用されるようになった現実がある。

このような現状を鑑み、このたび公益財団法人 NIRA 総合研究開発機構では、同一の質問項目によって構成される面接調査とモニター型インターネット調査<sup>2</sup>を全く同じタイミングで実施して国勢調査を国勢調査を含めた 3 者間での結果の比較を行い、その集計結果や補正の可能性をつまびらかにすることにより、インターネット調査を「正しく」利用するための基礎資料を提供することにした。具体的には、まず、国勢調査との共通項目については国勢調査を「正解<sup>3</sup>」（基準）として面接調査、インターネット調査それぞれとの比較を行い、次に、面接調査とインターネット調査の間でそれ以外の項目を比べる。さらに、これらの比較結果を用いて、インターネット調査データを面接調査データに合わせるかたちで補正を試みる。

## 2. 問題の所在

人々のインターネットの利用率は年々高まり、2019 年には 89.8%までに達した（総務省 2020）<sup>4</sup>。しかし、前述のとおり、それは必ずしもインターネット調査会社のモニターやそこから抽出された回答者の回答がインターネットユーザー、まして目標母集団である日本人全体と変わらなくなったことを意味しない。

日本学術会議社会学委員会 Web 調査の課題に関する検討分科会（2020）は、インターネットを用いた調査の問題点を、（1）サンプルと代表性の問題、（2）測定上の問題、（3）現在のインターネット環境から考えられる問題の 3 種類に整理している。

（1）サンプルと代表性の問題とは、インターネット調査は無作為抽出法を用いた調査ではないことに加え、モニターに登録しない人々の存在や早いもの順で回答を受け付けること、さらには調査協力にともなうポイントを稼ぐために調査への回答をこなしていくことを優先する者によって生じる、サンプルの偏りである。

（2）測定上の問題点とは、回答に利用するデバイスやブラウザ、利用する回線等によって調査票の見え方が異なったり、自記式調査に求められる準備が不足していたり、サティスファイサー（Satisficer、意味は後記）が含まれていたりするために生じる測定誤差である。

---

<sup>2</sup> 調査会社に事前に応募し、登録されたインターネットユーザー群（モニター、アクセスパネル）の中から調査対象者を募る、現在最もよく行われているタイプのインターネット調査。

<sup>3</sup> 厳密には、国勢調査にも無視できない程度の無回答が存在し（令和 2 年国勢調査の回答率は 8 割強）、市区町村や調査員による補記後も誤差がありうる。

<sup>4</sup> 20 歳以上の世帯主がいる世帯が調査対象。そのうち、最年少は 6 歳。

(3) 現在のインターネット環境から考えられる問題としては、高齢者のインターネット利用率は相対的に低いこと、インターネット利用者の中でもインターネット調査の回答者は2~3割以下と考えられること、気軽に調査に応じる人とインターネット利用に不安を感じる人との間で差がありうること、そして無作為抽出法を用いた調査における回収率のような、インターネット調査の質の善し悪しを判断する簡単な基準が見当たらないこと、が指摘されている。

これらの偏りを低減するための主な取り組みとしては、(1) 調査設計段階におけるサンプル割り付けによる補正、(2) 調査分析におけるウェイトを用いた補正、(3) サティスファイサー検出による補正の3つが挙げられる。

(1) 調査設計段階におけるサンプル割り付けによる補正とは、インターネット調査を行う前に、バイアスが生じやすい属性について国勢調査などと分布が等しくなるように、調査前に回収目標数を割り付けることを指す。割り付けに用いる属性は性別と年齢(層)が多いが、割り付けの条件を増やすと、必要数を満たせないセルが出て来てしまうという新たな問題が生じる。

(2) 調査分析におけるウェイトを用いた補正は、調査分析時点で補正(事後補正)する方法である。代表的なものについては次章で触れるが、セル・ウェイトイング法、レイキング法、傾向スコアを用いた重み付け推定法などがある。

(3) サティスファイサー(Satisficer・省力回答者)検出による補正とは、アンケート調査において質問文を注意深く読まずに労力を最小限化(satisfice)するような行動(回答)をとり(三浦・小林 2018)、調査データの質を毀損しかねない者を排除し、またはそうした行動を防止することである。サティスファイサーの検出および分析段階における補正方法として、三浦・小林(2015)は質問文をきちんと読んでいるかどうかを確認するトラップ質問(例、この質問では左から2番目の選択肢を回答してください)を取り入れることを推奨している。トラップ質問を混ぜると、データ分析段階においてサティスファイサーを特定し、取り除くことができるほか、調査中に回答者がトラップに気付き、注意して調査に取り組む効果も期待される。

### 3. 本プロジェクトの目的とリサーチクエスションの設定

さらに問題を複雑にしているのが、偏っているのはインターネット調査に限らないという事実である。「理想的」な確率標本抽出が可能な住民基本台帳による無作為抽出に

基づく訪問面接調査法・郵送調査法などの従来型調査は、多額の費用が掛かる割に（中瀬 2005）、個人情報保護に対する意識の高まりなどに伴って回収率が低下の一途をたどっており、非回答バイアスはもはや無視できない水準に達しつつある。他方、インターネット調査データには繰り返し述べている通りのバイアスが含まれており、そこから得られた知見を安易に一般化することは難しい。

ところが、全数調査である国勢調査とインターネット調査、無作為抽出に基づく調査を1対1で比較検討する研究はあっても、これらの3者をシステマティックに比較したものは意外に少ない（本田・本川（2005）、近年では小林（2019）など）。

例えば、世論調査では高年齢層と比べて若年層の回収率が低い点は広く知られており、無作為抽出されたサンプルを対象に実施された調査データを、事後的に国勢調査の年齢別人口構成に等しくなるようにウエイトバックする補正はしばしば行われている。しかし、無作為抽出に基づく調査が含むバイアスは年齢由来のものだけとは限らない。

あるいは、無作為抽出に基づく調査とインターネット調査で選挙の投票率を比べると、インターネット調査の方が実際の数値に近くなることも知られている。ただ、そうした現象はサティスファイサーによる可能性<sup>5</sup>が看過されており、インターネット調査データの方が投票参加の分析に有益であることを直ちには意味しない。

以上を踏まえ、本プロジェクトでは2つの問題に取り組むことにした。

第一に、無作為抽出に基づく面接調査とインターネット調査を同時に実施することにより、国勢調査結果を合わせて、各調査モードの回答者間にどのようなバイアスが生じているのかを観察する。リサーチクエスチョンとしては、次のようになる。

**RQ1：**目標母集団を18歳以上の日本国民としたとき、代表的な調査方法（無作為抽出面接調査・事前割り付けモニター型インターネット調査）から得られたデータにはそれぞれどのような（体系的な）バイアスが生じているのか。

本プロジェクトの調査では、基本的属性に加えて、調査間での差異の原因または結果となりそうな質問項目を投入する。具体的には、林（2016）で検討された性格5因子のほか、インターネット利用時間・利用目的である。

第二に、そのバイアスに対する補正の可能性についてである。すでに述べた通り、調査後のデータ補正については、様々な方法が採用されてきた。得られたデータに補正を行った場合、どの程度データの分布が変化するかを確認する必要がある。リサーチクエスチョンとしては、次のようになる。

---

<sup>5</sup> 調査に慣れたモニターは、どこに投票したか、その政党や候補者に投票した理由は何か、など「投票した」と答えた場合に続くであろう質問をスキップして回答を早く終わらせるため、本当は投票したのに「投票しなかった」と答えたかもしれないからである。

RQ2：無作為抽出面接調査・事前割り付けモニター型インターネット調査データに対し補正を施すと、各項目の分布はどの程度変化するのか。



## 第2章 研究方法

### 要旨

本章は、本プロジェクトが並行実施した面接調査とインターネット調査の概要および国勢調査を含めた比較結果を示す。面接調査とインターネット調査には、21大都市在住率、教育程度、住居の種類、5年前常住地、インターネット利用動向に差が見られる。ただし、面接調査データも、回答者の年齢、住居の種類、持ち家率、5年前常住地について国勢調査から乖離しており、必ずしも18歳以上の日本人の縮図とは言えない。後半では、主な調査データの補正方法を紹介した上で、本プロジェクトが採用した補正方法である、国勢調査サンプルデータと各種調査データを組み合わせた傾向スコアによるIPW法について説明する。

### 1. 調査概要と記述統計

#### 本プロジェクト調査データの概要

本プロジェクトでは、無作為抽出面接調査およびモニター型インターネット調査<sup>6</sup>という代表的な調査方法を踏襲しつつ、独自に設計した3種の世論調査（面接調査を1種類、インターネット調査を2種類）を、実施開始日を2021年4月2日に揃えて並行実施した<sup>7</sup>。各調査の概要は表2-1の通りである。

---

<sup>6</sup> インターネット調査においては、回答者の集め方として、自己参加型の「公募型」と従来型の標本抽出を行ったのちにインターネット上での回答を求める「非公募型」に整理されることもある（大隅 2006）。

<sup>7</sup> 調査の実施に当たっては、NIRA 総合研究開発機構倫理審査委員会の承認を得た（承認番号 20105\_2021R37\_02）。

表 2-1 本プロジェクトの調査概要

	面接調査	インターネット調査A	インターネット調査B
調査対象	全国の18歳以上の男女4,088人	調査会社A登録モニターのうち、 全国の18歳以上の男女1,500人	調査会社B登録モニターのうち、 全国の18歳以上の男女1,569人
有効回答数	1,124人（回収率27.5%）	1,500人 （サティスファイサー除外：1260人、84.0%）	1,569人 （サティスファイサー除外：1,426人、90.9%）
抽出方法	層化3段無作為抽出法 （電子住宅地図）	国勢調査（H27）に基づいた性別・ 年代（10歳刻み）割り当て 性別（男女）× 年代（18～29、30～39、40～49、50～59、 60～69、70～）	国勢調査（H27）に基づいた性別・ 年代（10歳刻み）割り当て 性別（男女）× 年代（18～29、30～39、40～49、50～59、 60～69、70～）
調査時期	2021年4月2日（金）～12日（月）	2021年4月2日（金）午前10時～6日（火）	2021年4月2日（金）午前10時～6日（火）
調査方法	調査員による訪問個別面接調査	Web画面上での個別記入法	Web画面上での個別記入法
調査項目（質問票）	共通		

面接調査は、電子住宅地図を用いた層化3段無作為抽出<sup>8</sup>により選ばれた調査対象者に、調査会社の面接員が対面で聴取したものである。ちなみに、本調査が行われた期間には、改正新型インフルエンザ等対策特措法に基づく緊急事態宣言は発令されていない。インターネット調査は、性別と年齢層の分布が国勢調査と同じになるように回収目標数を指定した上で、調査会社の登録モニターに回答を依頼した。

調査質問項目については、調査方法の違いに由来する細かな言い回しの違いを除いて、3種類全ての調査で共通とした。プロジェクトの目標に合わせ、国勢調査の調査項目を多く盛り込んだこと、そして面接調査とインターネット調査の比較を主眼として、差異を予測できる質問を採用したことが特徴である。以下、それぞれの質問項目の概略を述べる。具体的な質問文は付録を参照されたい。

## ■ 国勢調査との共通項目

### (1) 居住地域

面接調査では、訪問先の回答者が住む市区町村を調査会社が記録した。インターネット調査では、A社の回答者は居住地の郵便番号、B社の回答者は居住している市区町村を直接回答した。

### (2) 男女の別

### (3) 実年齢

<sup>8</sup> 全国の市町村を12ブロックに分類、さらにブロック内で市郡規模に分類し層化を行った上で、国勢調査の基本単位区から調査地点を抽出し（1段目の抽出）、さらに住宅地図データベースを用いて対象世帯を抽出し（2段目の抽出）、対象世帯の居住者のうち割り当てられた性別・年代に該当する人を調査対象とするもの（3段目の抽出）。

性別と年齢については、いずれの調査も、調査設計段階において、平成 27 年度（2015 年度）国勢調査にもとづき割り付けを行っている。それでもなお調査モードによって誤差が生じるかを確認するため質問した。

#### (4) 教育程度

教育程度は、国勢調査の質問形式と揃えて全ての調査で採用している。ただし、国勢調査では教育程度は大規模調査の年にしか質問されておらず、比較対象は平成 22 年（2010 年）調査であることに留意されたい。

#### (5) 配偶者の有無

国勢調査では、回答者の配偶者の有無について、「未婚」、「配偶者あり」、「死別」、「離別」という選択肢を設け質問している。本調査でも、同じ質問文を採用した。

#### (6) 労働力状態

#### (7) 勤めか自営かの別（従業上の地位）

#### (8) 本人の仕事の内容（産業）

国勢調査に倣い、「有職者かどうか」・（有職者に対し）「勤めか自営か」・「仕事の内容」を質問した。面接調査やインターネット調査に回答しやすい職業があるのかを確認することが、これらの質問を採用した目的である。なお、国勢調査では回答者は「勤め先・業主などの名称及び事業の内容」を記入し、それを調査者（国）がコーディングする。本調査では、平成 27 年度国勢調査匿名データの符号表を参考に、区分と具体例を示し、回答者に当てはまるものを選んでもらうかたちを採用した。

#### (9) 住居の種類

#### (10) 5 年前の常住地

国勢調査と同じく、回答者の居住状態を質問した。一般に面接調査では（特にオートロック付きの）集合住宅に住む調査対象者からの聴収が難しいことが指摘されている（中瀬 2005）。その点、任意の時間と場所で回答できるインターネット調査の方が、国勢調査の分布と近い結果が得られるかもしれない。5 年前の常住地は、国勢調査との比較を目的に採用した。

### ■ 面接調査とインターネット調査の共通項目

#### (1) サティスファイサー検出質問

2 つのインターネット調査においては、サティスファイサーを検出する質問を 1 問設定した。パーソナリティを計測するための同じ回答形式の質問が 10 個並んでいるうちの 6 番目と 7 番目の間に「この質問には、『おおよそ違うと思う』を選択してください。」

という指示を挿入したところ、A社では16.0%、B社では9.1%が指示されたものとは異なる肢を選択した。これらの者はサティスファイサーの疑いが強く、原則として分析から除外する<sup>9</sup>。

## (2) パーソナリティ (TIPI-J)

調査モードによって回答者のパーソナリティが異なるのか、というクエスチョンを検討するために設けた質問である。本調査では、10項目によりビッグ・ファイブ（外向性、協調性、勤勉性、神経症傾向、開放性）を測定する日本語版 Ten Item Personality Inventory (TIPI-J) を採用した（小塩ら 2012）。各項目は以下の通りである。

- ① 活発で外向的だと思う
- ② 他人に不満を持ち、もめごとを起こしやすいと思う
- ③ しっかりしていて、自分に厳しいと思う
- ④ 心配性で、うろたえやすいと思う
- ⑤ 新しいことが好きで、変わった考えを持つと思う
- ⑥ ひかえめで、おとなしいと思う
- ⑦ 人に気をつかう、やさしい人間だと思う
- ⑧ だらしなく、うっかりしていると思う
- ⑨ 冷静で、気分が安定していると思う
- ⑩ 発想力に欠けた、平凡な人間だと思う

回答者は上の10項目の質問に対して、それぞれ「1. 全く違うと思う」～「7. 強くそう思う」の7点尺度で答えた。

分析における尺度構成は、小塩ら（2012）に従い以下のように計算した。

外向性：項目①+（8-項目⑥）

→値が大きいほど、「外向性が高い」を意味する。

協調性：（8-項目②）+項目⑦

→値が大きいほど、「協調性が高い」を意味する。

勤勉性：項目③+（8-項目⑧）

→値が大きいほど、「勤勉性が高い」を意味する。

神経症傾向：項目④+（8-項目⑨）

→値が大きいほど、「神経症傾向が高い」を意味する。

---

<sup>9</sup> 本調査データに基づくサティスファイサーに関する詳細な分析は、大森（2021b）を参照されたい。

開放性：項目⑤＋（8-項目⑩）

→値が大きいほど、「開放性が高い」を意味する。

### (3) インターネットの利用時間・目的

面接調査とインターネット調査の回答者に違いが予測される大きな点として、インターネット利用度がある。大隅ら（2019）は、インターネット調査とそれ以外の調査の対象者間には「デジタル・デバイド」が存在すると主張する。

そこで、本調査においては、平日1日当たりのインターネット利用時間を問うことで、それぞれの回答者のインターネット利用度の違いを明らかにする。ただし、インターネット普及率は高まっており、利用時間のみでは差異をはっきりと捉えられないかもしれない。そこで、回答者のインターネット利用目的を合わせて聞くことにした。

利用時間については総務省情報通信政策研究所「令和2年度情報通信メディアの利用時間と情報行動に関する調査」、利用目的については総務省「平成29年度通信利用動向調査」の結果とも比較する。

### (4) 景況感

### (5) 生活満足感

景況感、生活満足感は、世論調査において問われることの多い項目である。インターネット調査の補正を行う際の目的変数・説明変数のいずれかに用いられうる項目として採用した。

### (6) 長期的党派性

### (7) 内閣支持

人々の政治意識に関する代表的項目である。マスメディア等が定期的を実施する世論調査と比較しやすいことから、調査項目として採用した。

## 2. 各調査データにおける質問回答分布の乖離

調査別の各項目集計結果は、大森（2021a）を参照されたい。本節では、面接調査データとインターネット調査データ<sup>10</sup>、そして国勢調査結果<sup>11</sup>を比較したときに、差が顕著であったもののみに触れておきたい。

---

<sup>10</sup> インターネット調査データの回答分布は、サティスファイサー除外後のものである。

<sup>11</sup> 比較対象は平成27年度国勢調査結果。教育程度のみ平成22年度国勢調査の結果。

## 面接調査とインターネット調査の差異

まず、基本的な社会属性については「21 大都市在住率」・「教育程度」・「住居の種類」・「5 年前常住地」に顕著な差が見られる（表 2-2）。

21 大都市に在住している者の割合は、面接調査で 25%程度である一方、インターネット調査データは A 社 38.4%、B 社 34.9%とともに 10 パーセントポイント以上高かった。よく行われている性別と年齢層に基づく事前割り付けのみでは、インターネット調査は大都市居住者がオーバーサンプルになることが分かる。

教育程度については、インターネット調査回答者の学歴は非常に高い。A 社・B 社の両方で、選択肢「大学・大学院」該当者の割合は 45%を超えており、面接調査（と国勢調査）の 20%台を大幅に上回る<sup>12</sup>。

住居の種類と 5 年前常住地については、面接調査の回答者において「持ち家」率と「現在と同じ住所」居住率が高い。

---

<sup>12</sup> 国勢調査では中途退学者はその前の学校を答えることとしている点を割り引いても、インターネット調査回答者が高学歴である点は変わらない。

表 2-2 基本的属性項目における調査間の差異

	面接調査	ネット調査A	ネット調査B	国勢調査
平均年齢(歳)	56.5	52.7	52.6	51.9
21大都市在住率(%)	25.3	38.4	34.9	27.7
教育程度：大学以上在籍・中退・卒業率(%)	24.1	47.2	45.7	21.9
住居の種類：「持ち家」率(%)	85.1	70.1	70.8	71.7
5年前常住地：「現在と同じ」率(%)	86.5	74.4	80.4	78.1
人数	1124	1260	1426	101173085

表 2-3 基本的属性項目以外における調査間の差異

	面接調査	ネット調査A	ネット調査B	総務省調査
				単位：%
1日のインターネット利用時間平均(分)	108.8	211.2	193.7	168.4
電子メールの送受信	56.0	83.1	83.2	76.6
天気予報	50.1	74.4	75.5	62.9
地図・交通情報の提供サービス	36.3	56.3	55.7	60.6
ニュースサイト	49.1	74.2	78.4	55.0
無料通話アプリやボイスチャット	29.5	36.0	34.6	52.9
ソーシャルメディア	20.0	36.1	38.1	52.2
動画投稿・共有サイト	23.3	31.2	29.8	50.7
ホームページ・ブログの開設・更新又は閲覧、書き込み	14.1	19.4	22.6	37.3
商品・サービスの購入・取引 (金融取引及びデジタルコンテンツ購入を除く)	29.6	56.8	49.0	36.9
辞書・事典サイト	18.3	21.4	21.0	33.8
オンラインゲーム	20.4	21.4	21.3	27.4
ラジオ、テレビ番組、映画などのオンデマンド	17.0	22.9	19.9	21.6
デジタルコンテンツの購入・取引	5.2	8.6	7.8	20.4
クイズ・懸賞応募、アンケート回答	6.0	45.9	53.4	19.8
金融取引	7.2	24.6	23.2	14.6
インターネットオークション	8.4	15.5	16.6	12.1
eラーニング	4.4	5.2	5.4	8.2
電子政府・電子自治体	2.7	4.8	4.4	5.2
人数	1124	1260	1426	1500

次に、属性以外の項目（表 2-3）を見ると、やはりインターネット利用関連の質問回答には顕著な差が見られる。「1日のインターネット利用時間（平日）」の平均値は、面

接調査とインターネット調査で 85~100 分の開きがある。総務省の情報通信メディアの利用時間と情報行動に関する調査には 70 歳以上のサンプルが含まれておらず、日本人の平均ネット利用時間の正確な値は分からないが、面接調査は過小、インターネット調査は過大である可能性がある。「インターネットの利用目的」についても、全体的にインターネット調査回答者では各項目の選択率が高い。特に「電子メールの送受信」「天気予報」「商品・サービスの購入・取引（金融取引及びデジタルコンテンツ購入を除く）」等が、面接調査と比較して高率になっている。

### 面接調査データの「ズレ」

インターネット調査データは面接調査データと比較して基本的な社会的属性をはじめ多くの調査項目についてバイアスを含んでいるという前項の知見は、これまでも多くの先行研究によっても指摘されてきたところである。

これに加えて、本プロジェクトの眼目は、基本的な社会的属性について全数調査である国勢調査と同一の質問項目を設けていた点にある。国勢調査データと比較した結果、無作為抽出を経たはずの面接調査データにも、表 2-2 にある通り、項目によって調査に回答しなかった者に起因するバイアスが生じていることが明らかになった。

まず、「年齢」については、国勢調査結果やインターネット調査回答者と比べて、面接調査データのみ平均が 4 歳程度高い<sup>13</sup>。

次に、「住居の種類」が注目される。先に面接調査データとインターネット調査間で乖離があると述べたが、実は「持ち家」率が国勢調査（71.7%）に近いのは、面接調査ではなく、インターネット調査である。面接調査データの「持ち家」率が 85.1%と 3 種類の調査の中で最も高く、国勢調査データとも 15%近い差がある。さらに、「5 年前常住地」も、面接調査では現在と同じ住所に住んでいたという回答が 86.5%と非常に高く、持ち家率の高さと平仄の合う結果である。

本プロジェクトにおける面接調査データの抽出方法は電子住宅地図を用いており、住民基本台帳や選挙人名簿を用いたものではないにしろ、無作為抽出に基づく面接調査の回答者も、もはや目標母集団の縮図と捉えてよいわけではなさそうだ。無作為抽出に基づく調査結果を「正解」に見立てて、インターネット調査データを補正することも、想定する母集団に近づくバイアスの改善を保証するものではないと言えよう。

---

<sup>13</sup> 男女別に平均年齢をみると、面接調査では特に男性の平均年齢が高い。国勢調査では 51.5 歳であるのに対し、面接調査では 57.9 歳であった。女性はあまり変わらない（面接調査：54.8 歳、国勢調査：54.3 歳）。



### 3. 調査データの補正方法

#### 代表的な補正方法

本プロジェクトが採用する調査の補正方法について説明する前に、多くの社会調査で用いられる事後補正の方法について簡単に述べておく。

#### セル・ウェイトニング、レイキング法

セル・ウェイトニングは基本的には1つの、レイキングは複数の共変量に着目して層別し、その共変量の周辺分布が母集団の周辺分布と等しくなるように各層に重みをつけて推定を行う方法である（星野 2010）。レイキング法では、まず特定の共変量に対して周辺分布が母集団の周辺分布と等しくなるように各セルの重みを計算し、さらにほかの共変量の周辺分布が母集団の集団分布と等しくなるように、と計算を繰り返していく。公的統計におけるウエイトは、これらの方法が用いられることが多い（高橋 2019）。しかし、これらの方法では、目標母集団の分布がセルごとに判明している必要があり、また、共変量として調整可能であるのは5~6変数に限られ、多くの共変量を考慮することは困難である（星野 2009）。

#### MRP 法

マルチレベルモデルを用いた MRP 法（Multilevel regression and poststratification）も注目を集めている方法の1つである。この方法では、得られたデータに対して階層回帰モデルを当てはめ、事後層化した各セルの母集団から重みつき調査推定値を算出していく（Gao et al. 2019; Gelman and Little 1997）。具体的には、全国の有権者を性別、年齢、学歴、職業、選挙区別に数千から十数万のグループに分け、それぞれの投票確率を推定した後、国勢調査のデータなどを使って、選挙区に各グループの有権者が何人いるかを計算し、大規模なコンピューターシミュレーションを行って候補者の得票率を予測するなどの例がある<sup>14</sup>。レイキング等と比較して多次元的に要因を考慮可能で、特に近年においては、メディア等が行う選挙予測等の世論調査でも活用されつつあるが、汎用的な補正方法として確立されているとは言い難い。

#### 傾向スコアを用いた IPW による重み付け法

ほかに多次元的な補正の方法として挙げられるのが、傾向スコアを用いた IPW（Inverse Probability Weighting）による重み付け法である。傾向スコアは、複数の共変量

---

<sup>14</sup> 「測れない数値を測る 世論調査の統計的モデリング」『中日新聞』、最終閲覧日 2022 年 3 月 25 日。 [https://static.chunichi.co.jp/chunichi/pages/feature/election/measuring\\_unmeasurable.html](https://static.chunichi.co.jp/chunichi/pages/feature/election/measuring_unmeasurable.html)

を一つの共変量に集約する方法であり、Rosenbaum and Rubin (1983) により考案された。傾向スコアは、例えばインターネット調査回答者であるか否かを従属変数とし、観察された共変量を投入した二項ロジスティック回帰分析で求められる。星野 (2007) は日本における適用例を多く挙げ、傾向スコアによる補正は多くの場合、上手くバイアスを改善すると主張する。

傾向スコアを用いた補正の場合、マッチングや層別解析も採用されるが、汎用性が高い点で上記の IPW 法に軍配が上がる。これは、求められた傾向スコアの逆数を用いて、各調査回答者に重みづける方法である。参照となるマイクロな調査データが必要であるが、より多くの共変量を投入できることが利点である。

### 本プロジェクトにおける補正方法：国勢調査調査票データを用いた検討

以上の検討を基に、本プロジェクトでは傾向スコアを用いた IPW 法を補正方法として採用することにした。直ちに逢着する課題は、母集団となる 18 歳以上の日本人について参照となるマイクロな調査データを得ることが難しい中、どのようにして傾向スコアを求めるか、である。これに対して本プロジェクトでは、公的統計のマイクロデータ利用、つまり国勢調査の調査票情報を用いた検討を行うことにした。平たく言えば、国勢調査の（ほぼ）ローデータを利用し、傾向スコアを求めるということになる。国勢調査のマイクロデータを「カバレッジや抽出のバイアスがない参照調査」（田島 2021）と捉え<sup>15</sup>、本プロジェクトの独自調査データと組み合わせることにより、傾向スコアを算出し、重み付けを行う。具体的な分析の手順について述べる。

**Step1** まず、国勢調査オンサイト利用で、平成 27 年国勢調査データ（ $n=122,000,000$  程度）の中から、調査対象となった日本人かつ 18 歳以上をソートした上で、 $n=3,000$  のサンプルを無作為抽出した。データの構成の都合上、平成 27 年国勢調査・18 歳以上日本人の集計データから、3,000 人を 47 都道府県別に割り付け、都道府県ごとに抽出したサンプルデータをマージした。以下では、これを「国勢調査サンプルデータ」と呼ぶ。

国勢調査データの全回答レコードを利用しなかった理由は、オンサイト内に設置された仮想 PC の性能上、1 億を超えるサンプルを扱うことが難しかったためである。作成した国勢調査サンプルデータの記述統計の情報は表 2-4 に示した通りで、公表されている国勢調査の全数集計<sup>16</sup>をほぼ再現できている。

**Step2** 国勢調査サンプルデータと、本プロジェクトが独自に行った面接・インターネット調査データをそれぞれマージする。マージされたデータを用いて、面接調査回答者か

---

<sup>15</sup> 正確を期すなら「カバレッジや抽出のバイアスが極めて少ない参照データ」。

<sup>16</sup> 数値は大森 (2021a) を参照。

国勢調査回答者かを示すモード変数を作成し、共通の質問項目を独立変数、モード変数を従属変数としたロジスティック回帰分析を行い、各サンプルにおける傾向スコアを求める。

**Step3** 算出された傾向スコアを用いて IPW 法による重み付けを行う。目的変数（例えば生活満足度）別にウエイトをかける前後の分布の変化を観察する。

表 2-4 国勢調査サンプルデータの記述統計

性別・年齢	
性別(女性%)	52.0
年齢平均(歳)	53.0
居住地域(%)	
21大都市在住	28.0
配偶者有無(%)	
未婚	23.9
配偶者あり	59.4
死別	8.8
離別	5.5
労働力状態(%)	
就業者	56.4
完全失業者	2.1
家事	14.1
通学	2.8
その他	18.8

従業上の地位(%)

---

雇用者（正規の職員・従業員）	28.3
雇用者（労働者派遣事業所の派遣社員）	1.3
雇用者（パート・アルバイト・その他）	14.6
役員	2.6
自営業主（雇人のある業主） /	
自営業主（雇人のない業主） / 家庭内職者	5.4
家族従業者	2.2

---

業種(%)

---

農業，林業，漁業	4.0
鉱業，採石業，砂利採取業，建設業	7.3
製造業，電気・ガス・熱供給，水道業	16.7
情報通信業	2.8
運輸業，郵便業	5.5
卸売業，小売業	15.5
金融業，保険業	2.5
不動産業，物品賃貸業	2.1
学術研究，専門・技術サービス業	2.9
宿泊業，飲食サービス業	6.4
生活関連サービス業，娯楽業	3.8
教育，学習支援業	4.0
医療，福祉	11.5
複合サービス事業/サービス業	
（他に分類されないもの）	6.6
公務（他に分類されるものを除く）	3.8
分類不能の産業	4.7

---

住居の種類(%)

持ち家	70.6
公営の借家	3.1
都市機構・公社の借家	1.2
民営の借家	18.7
給与住宅	1.7
間借り	0.6
会社等の独身寮・寄宿舍	0.6
その他	0.5

5年前常住地(%)

現住所	73.3
自市区町村内	8.6
自市内他区	1.2
県内他市区町村/他県/転入（国外から）	10.4

人数	3000
----	------

傾向スコアを求めるための基本モデルは、従属変数に面接またはインターネット調査回答者である確率、独立変数に調査で質問した基本的属性を投入したロジスティック回帰式である。

補正の目的変数は、第2章で説明した本プロジェクト調査の質問項目のうち「長期的党派性」、「内閣支持」、「生活満足度」、「景況感」「1日のインターネット利用時間」、「インターネット利用目的」である。先に断っておくと、これらは「正解」の数値があるわけではない。大森（2021a）で掲載した参考となるデータ——総務省「情報通信メディアの利用時間と情報行動に関する調査」、「NHK 世論調査」——も、何らかのサンプリングを経て行われた調査の結果であり、そこにはバイアスが含まれる以上、必ずしも真値とは言えない。ここでの趣旨は、計測不可能な「正解」に近づくかどうかではなく、補正によってそれぞれの調査の分布・平均値がどのように変化するかである。

国勢調査をはじめとする公的統計のマイクロデータ、特に調査票情報データが利用可能となったのは最近であり、まだ活用実績も少ない。本プロジェクトにおける各調査は、国勢調査の主要調査項目を包含する形で調査票を設計することにより、目標母集団（18歳以上の日本人）にほぼ相当するローデータを利用して調査データの補正を試みる先駆的な事例となろう。

## 第3章 分析と結果

### 要旨

第3章では各種調査の補正結果を示す。まず、面接調査に基づいてインターネット調査データを補正したときには、長期的党派性、内閣支持、生活満足度、景況感については、補正前後で大きな変化は認められなかった一方、1日当たりインターネット平均利用時間は補正により若干減少した。次に、国勢調査サンプルデータによって面接調査データを補正したところ、内閣支持率と不支持率が逆転したほか、インターネット平均利用時間が15分以上短くなる効果が見られた。国勢調査サンプルデータによってインターネット調査データを直接補正した場合には、各種の目標変数は小幅の変動にとどまった。

### 1. 従来型補正

各種調査の補正結果をみていこう。まず示すのは、(未補正の)面接調査を「正解」と見立て、それに合わせてインターネット調査データを補正した結果である。

インターネット調査会社Aのデータ(以下「Aデータ」と面接調査データをマージして、傾向スコアを求めるためのロジスティック回帰分析を行った結果が表3-1である。投入した共変量は、女性ダミー、年齢、教育程度、各パーソナリティ、既婚/死別ダミー、無職/家事/パート/自営ダミー、1次/2次産業従事ダミー、持ち家ダミー、5年前と常住地に変更なしダミー、各居住ブロックダミー(参照カテゴリは「北海道」)、21大都市在住ダミーである。従属変数は、Aデータへの回答者=1、面接調査データ回答者=0をとる。

各変数の係数をみると、女性ダミーや年齢・教育程度のほか、パーソナリティ項目(外向性・協調性・誠実性・神経症傾向)、既婚ダミー、死別ダミー、パートダミー、2次産業従事ダミー、持ち家ダミー、5年前常住地ダミー、関西ダミー、21大都市在住ダミーが有意水準5%を満たす。例えば、女性である、年齢が高い、最終学歴が高い、外向性や協調性が低い、誠実性や神経症的傾向が高い、配偶者と死別または既婚者ではない、パートタイマーではない、自宅を所有していない、5年前から住所が変わった、そして21大都市に住んでいる人ほど、(面接調査ではなく)A社のインターネット調査の回答者である確率が高く、各変数の記述統計におけるAデータと面接調査データの差異と

平仄が合っている<sup>17</sup>。

---

<sup>17</sup> 表 2-2・表 2-3 に示されているもの以外は大森（2021a）を参照されたい。

表 3-1 A データと面接調査データをマージしたロジスティック回帰分析結果:モデル1

	係数	標準誤差
(Intercept)	-0.77	0.53
女性	0.37	0.12 **
年齢	0.02	0.00 ***
教育	0.68	0.06 ***
外向性	-0.09	0.02 ***
協調性	-0.13	0.03 ***
誠実性	0.10	0.02 ***
神経症傾向	0.10	0.02 ***
開放性	0.03	0.02
既婚	-0.38	0.13 **
死別	-1.29	0.23 ***
無職	-0.27	0.17
家事	0.20	0.17
パート	-0.40	0.15 **
自営	-0.12	0.21
1次産業	-0.33	0.43
2次産業	-0.33	0.16 *
持ち家	-0.78	0.13 ***
常住	-0.58	0.14 ***
東北	0.22	0.30
関東	0.33	0.25
甲信越	-0.15	0.34
北陸	-0.17	0.41
東海	0.49	0.27
関西	0.65	0.26 *
中国	-0.30	0.31
四国	0.31	0.36
九州	-0.07	0.28
21大都市	0.33	0.11 **
人数	2282	

\*\*\*p<0.001 \*\*p<0.01 \*p<0.05



次に、傾向スコアを用いてサンプルに重み付けを行った上で、A データの各項目の分布を算出した<sup>18</sup> (表 3-2)。長期的党派性、内閣支持、生活満足度、景況感については、補正前と補正後で大きな変化は認められなかった。一方、インターネット利用時間については変化が見られ、重み付け前が 1 日平均 210.82 分であったのに対し、重み付け後は 204.20 分と、6 分以上短くなった。

表 3-2 モデル 1 を用いた A データの補正前・補正後の分布と平均 (面接調査データに合わせた補正)

長期的党派性	補正後	補正前
自由民主党	29.2	28.7
公明党	2.5	2.2
立憲民主党	6.0	6.0
日本共産党	2.9	3.2
日本維新の会	5.9	6.3
国民民主党	0.6	0.6
社会民主党	0.5	0.5
れいわ新撰組	0.9	1.0
NHK受信料を支払わない方法を教える党	0.8	0.9
その他の政党	0.4	0.4
そのような政党はない	32.7	31.7
わからない・答えたくない	17.9	18.5

内閣支持	補正後	補正前
支持しない	44.1	44.4
支持する	23.8	23.6
わからない・答えたくない	32.2	32.0

<sup>18</sup> 分析には、統計ソフト R の GDAtools、Stats パッケージを用いた。

生活満足度	補正後	補正前
不満だ	5.6	5.8
やや不満だ	16.6	16.1
どちらともいえない	21.5	20.2
まあ満足している	46.8	48.0
満足している	9.5	9.5

景況感	補正後	補正前
かなり悪い	2.8	3.2
やや悪い	15.6	15.3
変わらない	69.4	68.0
やや良い	8.9	10.1
かなり良い	3.2	3.3

1日のインターネット利用時間	補正後	補正前
平均（分）	204.20	210.82

インターネット利用目的	補正後	補正前
ネット1	3.62	3.67
ネット2	0.52	0.53
ネット3	1.43	1.47
ネット4	0.46	0.45

インターネット調査はインターネットを使わなければ回答できないのに対し、面接調査はインターネットユーザーである必要はない。そこで、表 3-3 は、インターネット利用に関わる質問項目を補正変数に加えた結果である。インターネット利用時間は対数変換した値、ネット利用目的については、ネット 1 は「メールの送受信」、「天気予報」といった情報収集を、ネット 2 は「金融取引」、「電子政府・電子自治体」といった商取引やオンラインでの行政機関の利用、ネット 3 は「動画投稿サイト」や「ゲーム」、「SNS」といった娯楽、そしてネット 4 は「アンケート回答」を示している。それぞれの指標作成の詳細は付録を参照されたい。

ロジスティック回帰分析（従属変数は A データへの回答者=1、モデル 2）の結果を見ると、インターネット利用関連の変数では、インターネット利用時間、ネット 3（娯楽）、

ネット4（アンケート回答）が従属変数に対して有意なプラスの効果を持つ。ネット利用時間が長いほど、インターネット利用目的が娯楽ではない、そしてアンケート回答目的でインターネットを利用する人ほど、Aデータの回答者である確率が高まる。

表 3-3 A データと面接調査データをマージしたロジスティック回帰分析結果:モデル2

	係数	標準誤差	
(Intercept)	-4.19	0.70	***
女性	0.36	0.13	**
年齢	0.03	0.01	***
教育	0.52	0.07	***
外向性	-0.07	0.02	**
協調性	-0.15	0.03	***
誠実性	0.08	0.03	**
神経症傾向	0.11	0.03	***
開放性	-0.03	0.03	
ネット利用時間 (対数)	0.77	0.06	***
net1(メール・情報収集など)	-0.01	0.04	
net2(金融取引・電子政府など商取引)	0.17	0.09	
net3(動画・ゲーム・SNSなど娯楽)	-0.36	0.06	***
net4(アンケート回答)	2.25	0.17	***
既婚	-0.48	0.15	**
死別	-0.72	0.30	*
無職	-0.08	0.16	
パート	-0.51	0.18	**
自営	-0.39	0.25	
1次産業	0.08	0.53	
2次産業	-0.45	0.18	*
持ち家	-0.72	0.16	***
常住	-0.62	0.16	***
東北	0.64	0.36	
関東	0.41	0.28	
甲信越	-0.08	0.40	
北陸	0.02	0.49	
東海	0.71	0.31	*
関西	0.82	0.30	**
中国	-0.35	0.38	
四国	0.58	0.42	
九州	0.20	0.33	
21大都市	0.25	0.13	
人数	2282		

\*\*\* $p < 0.001$  \*\* $p < 0.01$  \* $p < 0.05$

ただし、モデル2で求めた傾向スコアを用いた補正の結果、表3-4のとおり、長期的党派性、内閣支持、生活満足度、景況感を大きく変化させるものではない<sup>19</sup>。

表3-4 モデル2を用いたAデータの補正前・補正後の分布と平均（面接調査データに合わせた補正）

長期的党派性	補正後	補正前
自由民主党	28.5	28.7
公明党	2.5	2.2
立憲民主党	5.4	6.0
日本共産党	3.0	3.2
日本維新の会	5.7	6.3
国民民主党	0.6	0.6
社会民主党	0.4	0.5
れいわ新撰組	0.8	1.0
NHK受信料を支払わない方法を教える党	0.6	0.9
その他の政党	0.4	0.4
そのような政党はない	31.6	31.7
わからない・答えたくない	20.4	18.5

内閣支持	補正後	補正前
支持しない	41.5	44.4
支持する	23.5	23.6
わからない・答えたくない	34.9	31.9

<sup>19</sup> このモデルにおいてはインターネット利用関連の変数を傾向スコアの算出に用いているため、インターネット利用関連の変数は補正の目的変数とはならない。なお、Aデータを調査会社Bのデータ（以下「Bデータ」）に置き換えて分析を行った結果も、基本的に同様であった。

生活満足度	補正後	補正前
不満だ	5.6	5.8
やや不満だ	16.6	16.1
どちらともいえない	21.5	20.2
まあ満足している	46.8	48.0
満足している	9.5	9.5

景況感	補正後	補正前
かなり悪い	2.6	3.2
やや悪い	16.0	15.3
変わらない	67.7	68.0
やや良い	9.9	10.1
かなり良い	3.7	3.3

## 2. 国勢調査の調査票データを用いた直接補正

本節で行うのは、前章で説明した国勢調査サンプルデータを用いた補正である<sup>20</sup>。まず、面接調査データについて、国勢調査サンプルデータとマージし、傾向スコアを求めたのち、目標変数について重み付けした結果を示す。次に同様の分析を、各インターネット調査データ（A データ・B データ）と国勢調査サンプルデータの組み合わせで行った結果を示す。

### 国勢調査を用いた面接調査の補正

面接調査データの補正に向け、傾向スコアを算出するために行ったロジスティック回帰分析（表 3-5）の従属変数は、面接調査回答者=1、国勢調査サンプルデータ回答者=0をとるダミー変数である。補正に投入した変数は女性ダミー、年齢、既婚／死別ダミー、無職／家事／パート／自営ダミー、1 次／2 次産業従事ダミー、持ち家ダミー、5 年前と常住地に変更なしダミー、各居住ブロックダミー（参照カテゴリは「北海道」）、21 大都市在住ダミーである。係数が有意水準 5%を満たした変数は、女性ダミー、既婚ダミー、死別ダミー、無職ダミー、家事ダミー、パートダミー、持ち家ダミーであった。すなわち男性、既婚者、配偶者と死別した人、無職ではない人、家事労働者、パートタイ

<sup>20</sup> データは統計法に基づき、独立行政法人統計センターから「国勢調査」（総務省）の調査票情報の提供を受け、分析者が独自に作成・加工したものであり、総務省が作成・公表している統計等とは異なる。

マー、自宅を所有している人であるほど、面接調査の回答者である確率が高いことを意味する。

表 3-5 面接調査データと国勢調査サンプルデータをマージしたロジスティック回帰分析結果

	係数	標準誤差
intercept	-1.65	0.24 ***
女性	-0.31	0.09 ***
年齢	0.00	0.00
既婚	0.28	0.10 **
死別	0.60	0.17 ***
無職	-0.24	0.12 *
家事	0.90	0.13 ***
パート	0.40	0.12 ***
自営	0.16	0.16
1次産業	-0.35	0.26
2次産業	0.03	0.11
持ち家	0.47	0.11 ***
常住	0.19	0.12
東北	-0.13	0.23
関東	-0.05	0.19
甲信越	-0.03	0.25
北陸	-0.18	0.30
東海	-0.03	0.21
関西	-0.27	0.20
中国	-0.10	0.23
四国	-0.04	0.28
九州	-0.15	0.21
21大都市	-0.04	0.09
人数	3743	

\*\*\*p<0.001 \*\*p<0.01 \*p<0.05

傾向スコアを用いた補正の結果は表 3-6 のとおりである。内閣支持について、補正前よりも約 3 パーセントポイント支持率が低く、不支持率が高くなり、支持と不支持が逆転した (DK・NA 率はあまり変わらない) 長期的党派性、生活満足度、景況感について

は、大きな分布の変化は見られなかった。

さらに注目されるのは、インターネット利用に関する大きな補正効果である。1日当たりのインターネット利用時間は、補正前が平均 108.81 分であったのに対し、補正後の平均値は 123.87 分と 15 分以上長くなった。総務省の情報通信メディアの利用時間と情報行動に関する調査（令和 2 年度）における 1 日当たりインターネット利用時間は平均 168.4 分であったが、同調査対象にはインターネット利用頻度が相対的に低いはずの 70 歳以上の高齢者が含まれていない一方、デジタルネイティブである 13～17 歳が入っている点を勘案すれば、18 歳以上の日本人全体（≒成人、有権者）の値としては恐らく過大であろう。この点、日本人成人による 1 日当たりインターネット利用は平均 2 時間という本プロジェクトの補正結果は、それなりにもっともらしく思われる。なお、インターネット利用目的についても、表 3-6 に示されているとおり、補正の結果、いずれの目的についても利用度を示す点数が高まっている。

表 3-6 面接調査データの補正前・補正後の分布と平均（国勢調査サンプルデータに合わせた補正）

長期的党派性	補正後	補正前
自由民主党	32.2	33.2
公明党	3.9	4.2
立憲民主党	5.8	6.2
日本共産党	1.7	1.7
日本維新の会	3.1	3.2
国民民主党	0.8	0.7
社会民主党	0.7	0.7
れいわ新撰組	0.4	0.4
NHK受信料を支払わない方法を教える党	0.6	0.6
その他の政党	0.1	0.1
そのような政党はない	40.2	37.9
わからない・答えたくない	10.6	11.1

内閣支持	補正後	補正前
支持しない	36.1	33.6
支持する	33.3	36.7
わからない・答えたくない	30.7	29.7



生活満足度	補正後	補正前
不満だ	2.5	2.3
やや不満だ	11.5	10.8
どちらともいえない	21.7	22.3
まあ満足している	49.5	49.6
満足している	15.0	14.7

景況感	補正後	補正前
かなり悪い	1.8	1.9
やや悪い	15.9	16.2
変わらない	73.9	74.1
やや良い	5.2	4.8
かなり良い	3.2	2.8

1日のインターネット利用時間	補正後	補正前
平均（分）	123.87	108.81

インターネット利用目的	補正後	補正前
ネット1	2.56	2.39
ネット2	0.25	0.23
ネット3	1.28	1.10
ネット4	0.07	0.06

### 国勢調査を用いたインターネット調査の補正

最後に、国勢調査サンプルデータを用いてインターネット調査データを補正する。傾向スコア算出のため行ったロジスティック回帰分析（表 3-7）の従属変数は、A データ・B データいずれとの組み合わせも、インターネット調査回答者であれば 1、国勢調査サンプルデータ回答者であれば 0 をとるダミー変数である。補正に投入した変数は前項の面接調査データ補正時と同じである。推定の結果、有意水準 5%を満した変数は、A データとのマージデータでは実年齢、死別ダミー、無職ダミー、家事ダミー、1 次産業従事ダミー、2 次産業従事ダミー、常住地ダミー、21 大都市ダミー、B データとのマージデータでは、女性ダミー、死別ダミー、家事ダミー、パートダミー、自営ダミー、1 次産業従事ダミー、2 次産業従事ダミー、持ち家ダミー、東北ダミー、中国ダミー、九

州ダミー、21大都市ダミーであった。Aデータ・Bデータを通じて、配偶者と死別した人、家事のみへの従事者、1・2次産業従事者、そして21大都市に住んでいない人などが、インターネット調査の回答者であるよりも、国勢調査サンプルデータに含まれている確率が高い。両データとも、21大都市ダミーについて有意な関連を持っており、インターネット調査回答者には21大都市在住者が多いことが再確認できる。

表3-7 インターネット調査データと国勢調査サンプルデータをマージしたロジスティック回帰分析結果

	Aデータ		Bデータ	
	係数	標準誤差	係数	標準誤差
intercept	-0.87	0.22 ***	-0.42	0.20 *
女性	-0.06	0.08	-0.17	0.08 *
年齢	0.01	0.00 **	0.00	0.00
既婚	-0.09	0.09	-0.17	0.09
死別	-0.79	0.19 ***	-0.68	0.17 ***
無職	-0.43	0.12 ***	-0.18	0.11
家事	0.88	0.12 ***	0.94	0.12 ***
パート	-0.14	0.11	0.29	0.10 **
自営	0.25	0.16	0.38	0.15 *
1次産業	-1.05	0.34 **	-1.53	0.41 ***
2次産業	-0.36	0.12 **	-0.30	0.11 **
持ち家	-0.16	0.09	-0.28	0.09 **
常住	-0.31	0.10 **	0.13	0.10
東北	-0.03	0.23	-0.44	0.21 *
関東	0.22	0.19	-0.06	0.16
甲信越	-0.24	0.27	-0.41	0.24
北陸	-0.22	0.33	-0.29	0.28
東海	0.28	0.20	-0.23	0.18
関西	0.28	0.20	-0.07	0.17
中国	-0.39	0.25	-0.43	0.21 *
四国	0.14	0.28	-0.40	0.27
九州	-0.36	0.22	-0.72	0.20 ***
21大都市	0.42	0.08 ***	0.28	0.07 ***
人数	3894		4066	

\*\*\*p<0.001 \*\*p<0.01 \*p<0.05

国勢調査データにより直接インターネット調査を補正したときの効果は、小幅にとどまった。すなわち、長期的党派性、内閣支持、生活満足度、景況感のいずれについても、補正による変動幅はほぼ2パーセントポイントの範囲内であった（表3-8）。

インターネット利用時間については若干の変化が見られ、Aデータでは、重み付け前は平均利用時間が211.2分であるのに対し、重み付け後の値は205分に、Bデータでも、重み付け前の193.38分から重み付け後は190.56分に、それぞれ短くなっている。

表3-8 インターネット調査データの補正前・補正後の分布と平均（国勢調査サンプルデータに合わせた補正）

内閣支持	データA補正後	データA補正前	データB補正後	データB補正前
支持しない	44.0	43.0	44.8	46.9
支持する	23.3	22.9	22.3	21.0
わからない・答えたくない	32.7	34.1	30.9	32.1

長期的党派性	データA補正後	データA補正前	データB補正後	データB補正前
自由民主党	29.2	27.5	25.5	24.1
公明党	2.4	2.3	1.7	2.0
立憲民主党	6.1	5.6	8.8	8.5
日本共産党	3.1	2.9	4.0	4.1
日本維新の会	5.9	5.8	5.2	5.0
国民民主党	0.1	0.7	0.7	0.7
社会民主党	0.7	0.5	0.9	0.8
れいわ新撰組	0.9	1.1	1.1	1.1
NHK受信料を支払わない方法を教える党	0.8	1.0	1.3	1.2
その他の政党	0.3	0.3	0.2	0.2
そのような政党はない	31.9	30.7	37.9	37.7
わからない・答えたくない	18.1	21.4	12.8	14.7

生活満足度	データA補正後	データA補正前	データB補正後	データB補正前
不満だ	5.9	5.6	11.3	11.0
やや不満だ	16.3	15.3	20.4	20.1
どちらともいえない	21.2	25.3	23.2	26.7
まあ満足している	47.2	44.7	39.8	37.5
満足している	9.4	9.0	5.41	4.8

景況感	データA補正後	データA補正前	データB補正後	データB補正前
かなり悪い	3.0	3.2	4.9	5.0
やや悪い	16.2	14.3	19.1	18.4
変わらない	68.5	68.7	66.2	66.5
やや良い	9.2	10.5	7.8	8.3
かなり良い	3.0	3.3	1.93	1.8

1日のインターネット利用時間 平均 (分)	データA補正後	データA補正前	データB補正後	データB補正前
	205.00	211.20	190.56	193.38

インターネット利用目的	データA補正後	データA補正前	データB補正後	データB補正前
ネット1	3.62	3.66	3.57	3.63
ネット2	0.52	0.53	0.52	0.52
ネット3	1.45	1.47	1.40	1.42
ネット4	0.46	0.45	0.53	0.53

## 第4章 結論と暫定的対策

### 要旨

本プロジェクトの調査により、事前割り付けモニター型インターネット調査には、割り付けに用いられていない基本的属性において国勢調査の集計結果からバイアスが生じていることのほか、インターネット利用に関しても、面接調査よりも利用時間が長く、利用目的の選択が多岐にわたる傾向が明らかになった。他方、無作為抽出に基づく面接調査にも、居住の形態や常住地について、国勢調査からの乖離が見られた。以上を踏まえ、本章では今後の社会調査に向けたいくつかの暫定的対処方法を提案する。各種調査の補正効果には、目標変数により近く、真値を入手可能な補正変数の有無が影響するものと思われる。

本プロジェクトでは、インターネット調査の利用がますます高まる現在において、インターネット調査のサンプルの特性と、その補正の可能性について探るべく、独自の調査を行った。

第1章で提示したリサーチクエスチョンに即して、得られた知見をまとめる。

1つ目のクエスチョンは「RQ1：目標母集団を18歳以上の日本国民としたとき、代表的な調査方法（無作為抽出面接調査・事前割り付けモニター型インターネット調査）から得られたデータにはそれぞれどのような（体系的な）バイアスが生じているのか」であった。

本プロジェクトの分析結果からは、事前割り付けモニター型インターネット調査には、割り付けに用いられていない基本的属性（例、居住地域、教育程度）において目標母集団の悉皆調査である国勢調査の集計結果と比較してバイアスが生じているほか、インターネット利用に関する項目についても、面接調査よりも利用時間が長く、利用目的の選択が多岐にわたるといった傾向が見られた。日本人の大半がインターネットを利用するようになったとしても、モニター型インターネット調査でアクセス可能な人々は、依然として日本人全体の映し鏡とは言えない。

他方、無作為抽出に基づく面接調査という従来型調査ももはや安心できるものではないことも示唆された。従来型調査における回収率の低下に伴う無回答バイアスは、決して無視できない水準に達している。本プロジェクトの面接調査では、居住の形態や常住地といった質問項目について、国勢調査とかけ離れた結果が示されていた。従来型調査を「正解」に見立ててインターネット調査データ補正をすることも、どれほどの意味を

持ちうるか、慎重にならざるを得ないのである。

2つ目のクエスチョンは「RQ2：無作為抽出面接調査・事前割り付けモニター型インターネット調査データに対し補正を施すと、各項目の分布はどの程度変化するのか。」であった。

この点に関しては、補正できる場合とできない場合があるというのが、本プロジェクトの結論である。本プロジェクトが設定した目標変数のうち、内閣支持や政党支持、景況感や生活満足度に関しては、国勢調査を用いた面接調査の補正——いずれも傾向スコアによる IPW 補正——で内閣支持率に変化が見られた以外は、それほど大きな補正効果は観察できなかった。例えば内閣支持や政党支持に対する前回選挙での投票政党／各党得票率のように、目標変数により近く、真値を入手できる補正変数があれば、より大きな補正効果が期待できたであろう。

他方、インターネット利用に関わる項目では、特に国勢調査を用いた面接調査の補正において、1日当たりの平均インターネット利用時間を大きく変化させる効果が得られた。そこには、年齢が高い、昔から同じ家に住んでいる、自宅を所有している（インターネット環境を自ら整えなければならない）などのインターネット利用に不利になりそうな要因を備えている人ほど面接調査には回答しやすかった、という背景が考えられる。トートロジーにならない範囲で目標変数に近い補正変数を得られるかどうか、補正効果を左右すると言えそうだ。

以上の知見を踏まえると、無作為抽出に基づく調査やインターネット調査を実施する上では、次の諸点を考慮する必要がある。

第一に、特に面接調査においては、居住形態や常住地の変化といった変数を用いたレイキング補正が一定程度有効である。面接調査には、持ち家に長年住んでいる「安定した居住環境にある人」が回答しやすい。それらの要因を調査対象者の無作為抽出時に考慮することは現実的ではなく、本プロジェクトが実施したような国勢調査のマイクロデータを用いた事後補正も、一般的に行える手法とは言えない。国勢調査のうちズレが生じやすい属性を面接・郵送調査やインターネット調査の質問項目に含めておき、事後的に分布を確認した上で、レイキング等で補正する方法が次善策となろう。

第二に、インターネット調査では、サティスファイサーを除去した後の回答者の年齢・性別・居住地域の構成比が目標母集団を反映するように、目標回収数を多めに設定すべきである。年齢と性別だけによる回収目標数の割り付けでは、都市部偏重になる。また、インターネット調査の回答者には、調査会社によりモニター管理がされているにもかかわらず、サティスファイサーがある程度——多いものでは15%以上——存在した。事前にデモグラフィック属性で回収目標数の割り付けを行っても、サティスファイサー除去後の属性分布は歪む。サティスファイサーを検出する質問を投入してサティスファ

イサー除去した後も、回答者が分析に必要なサイズを保ち、かつ目標母集団と変わらぬ属性分布となるよう、あらかじめ多めに目標回収数を設定することが、現実的な方策となろう。

第三に、調査目的となる変数が、(モニターに登録するような) インターネットユーザーであることによって影響されそうな場合は、インターネット利用時間などを対象とした適切な補正を行うべきである。本プロジェクトの結果からは、インターネットを頻繁に利用するモニターに特有の傾向は、インターネット利用時間の長さや利用目的の選択に反映されることが分かった。例えば DX 導入率、テレワークの利用率、メタバース利用経験などをインターネット調査で計測したとき、調査結果を日本人全体の意識と称するためには、そもそもインターネットユーザーが回答者である点を割り引く補正を行う必要がある。

そして第四に、1 回限りのインターネット調査で「知りたいもの」を計測できると過度の期待をしないことが重要である。2021 年衆院選では、朝日新聞がインターネット調査を併用した選挙情勢報道を行い、かつ、かなりの精度で予想的中させたことが話題を呼んだ。注意すべきは、同社はインターネット調査での「小選挙区でどの候補者に投票するか」の数字をそのまま用いるのではなく、比例代表での投票予定政党について電話調査を併用して補正を行い、かつ、過去の選挙における予測と実際の選挙結果のズレの傾向を把握した上で、今回の選挙結果を推測した点である<sup>21</sup>。予測したい数字(実際の得票率)とインターネット調査で計測した数字(投票意向)の間には埋めきれない乖離があることを認めた上で、経験則としての乖離傾向を予測に生かしたのである。選挙結果のような「正解」が存在しない場合にもいっても、同じ方法に基づくインターネット調査で、同じ質問を繰り返し行えば、世論調査ならぬ「世論観測<sup>22</sup>」として、前回調査との差分から現実を推論できるようになる。

もちろん、これらは暫定的な対処方法に過ぎない。社会調査の設計は常に回答者の負担と調査者の関心がせめぎあい、補正のために必要な変数(共変量)をすべて調査項目に含めることは不可能である。調査者が調査目的や目標母集団の性質をよく吟味し、調査設計を行うことが望まれる。

加えて今後は、国勢調査などの公的統計の調査票情報を活用した補正方法など、より厳密かつ汎用性のある調査方法のさらなる開発も望まれるところである。本プロジェクトが、調査環境の激変に耐え、人々の属性や知りたい「態度」を正しく掬い上げることのできる社会調査を目指した研究の進展に向けた礎石となりうるならば幸いである。

---

<sup>21</sup> 「(世論調査のトリセツ) 選挙予測でネット導入、偏り補正」朝日新聞デジタル、2021 年 11 月 12 日。 [https://www.asahi.com/articles/DA3S15109923.html?iref=pc\\_rensai\\_long\\_478\\_article](https://www.asahi.com/articles/DA3S15109923.html?iref=pc_rensai_long_478_article)

<sup>22</sup> 萩原 (2015)

## 参考文献

- 大隅昇(2002)「インターネット調査の適用可能性と限界—データ科学の視点からの考察—」『行動計量学』29(1),pp.20-44.
- (2006)「インターネット調査の抱える課題と今後の展開」『エストレーラ』143(2), pp2-11.
- 大隅昇・嶋真紀子・井田潤治・小野裕亮(2019)『ウェブ調査の科学』朝倉書店.
- 大森翔子 (2021a)「インターネット調査のサンプル特性：国勢調査・面接調査との比較」NIRA ワーキングペーパーNo.1.
- (2021b)「インターネット調査における省力回答者に関する一考察」NIRA 政策研究ノート vol.3.
- 小塩真司・阿部晋吾、・カトローニ・ピノ (2012)「日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み」『パーソナリティ研究』21,pp.40-52.
- 小林良彰 (2019)「社会意識に関する異なる調査方法比較：インターネット調査と面接調査の比較検討」『法学研究』92(4),pp.1-21.
- 総務省 (2020)『令和2年 通信利用動向調査報告書 (世帯編)』  
[https://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR202000\\_001.pdf](https://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR202000_001.pdf) (最終閲覧日：2022年3月27日)
- 田島恵美 (2021)「混合研究法におけるウェブ調査利用の可能性と限界」『大正大学研究紀要』106,pp.93-106.
- 谷口将紀・谷口尚子 (2008)「インターネット調査の可能性—東京大学・朝日新聞社共同世論調査の比較—」『日本政治研究』5(1・2 合併),pp.222-233.
- 中瀬剛丸 (2005)「面接調査の現状と課題：調査実施の立場から」『放送研究と調査』2005年12月号,pp.66-73.
- 日本学術会議社会学委員会 Web 調査の課題に関する検討分科会 (2020)「提言 Web 調査の有効な学術的活用を目指して」日本学術会議, (2022年3月11日)
- 日本マーケティング・リサーチ協会 (2021)「第46回経營業務実態調査」
- 三輪哲・石田賢示・下瀬川陽 (2020)「社会科学におけるインターネット調査の可能性と課題」『社会学評論』71(1), pp.29-49.
- 萩原牧子(2009)「インターネットモニター調査はどのように偏っているのか—従来型調査手法に代替する調査手法の模索—」『Works Review』Vol.4,pp.8-19.
- 萩原雅之 (2015)「インターネット調査による世論観測の試み—「空気」の変化を詳細・迅速に捉えるための発想と実践—」『政策と調査』9,pp.51-58.
- 星野崇宏 (2007)「インターネット調査に対する共変量調整法のマーケティングリサーチへの適用と調整効果の再現性の検討」『行動計量学』34(1),pp.33-48.
- (2009)『調査観察データの統計科学』岩波書店.
- (2010)「調査不能がある場合の標本調査におけるセミパラメトリック推定と感度分析：日本人の国民性調査データへの適用」『統計数理』58(1),pp3-23.
- 本多則恵・本川明 (2005)「インターネット調査は社会調査に利用できるか—実験調査による検証結果—」『労働政策研究報告書』No.17,労働政策研究・研修機構.



- 三浦麻子・小林哲郎 (2015) 「オンライン調査モニタの Satisfice に関する実験的研究」『社会心理学研究』  
31,pp.1-12.
- (2018) 「オンライン調査における努力の最小限化が回答行動に及ぼす影響」『行動計量学』45(1),  
pp1-11.
- 林明明 (2016) 「インターネット調査と従来型紙面調査による調査結果に違いはあるのか」『電気通信普及  
財団 研究調査助成報告書』No.31,pp.1-8.
- Gao, Y., Kennedy, L., Simpson, D., and Gelman, A. (2021). Improving multilevel regression and poststratification  
with structured priors. *Bayesian Analysis*, 16(3), 719-744.
- Gelman, A., and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression.
- Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for  
causal effects. *Biometrika*, 70(1), 41-55.

## 付録

### 付録 A 各調査における質問と選択肢

本プロジェクトでは各調査で共通の質問を行っているが、面接調査においては選択肢の多い質問のみ、調査員が選択肢の記載された用紙を回答者に提示し、答えさせている。その質問については◎印を付した。

また、ほとんどの質問には、「わからない・答えない」の選択肢を設けていない。インターネット調査では、いずれかの選択肢で回答をしないと終了できない仕組みになっているが、面接調査では、口頭で回答を拒否され、「わからない・答えない」とコーディングされることもある。面接調査のみ「わからない・答えない」とコーディングされることがあった質問については、カギ括弧 (【】) で示した。

問1 あなたのお住まいの地域をお教えてください。

都道府県 ( ) 市区町村 ( )

問2 あなたの性別をお答えください。

1. 男
2. 女

問3 現在の年齢は、おいくつですか。

□□歳

問4 あなたが最後に在籍した（または現在、在籍している）学校はこの中のどれにあたりますか。

1. (新) 中学 / (旧) 小・高小
2. (新) 高校 / (旧) 中学
3. (新) 短大・高専 / (旧) 高校・専門
4. (新) 大学・大学院 / (旧) 大学・大学院

問5◎ 次の1から11まで（面接調査では1から10まで）のことばがあなた自身にどのくらい当てはまるかについて、もっとも適切なものをそれぞれ1つずつ選んでください。

私は自分自身のことを

1. 全く違うと思う
2. おおよそ違うと思う
3. 少し違うと思う
4. どちらでも

ない 5. 少しそう思う 6. まあそう思う 7. 強くそう思う 【8. わからない】

1. 活発で外向的だと思う
2. 他人に不満を持ち、もめごとを起こしやすいと思う
3. しっかりしていて、自分に厳しいと思う
4. 心配性で、うろたえやすいと思う
5. 新しいことが好きで、変わった考えを持つと思う
6. ひかえめで、おとなしいと思う
7. この質問には、「おおよそ違うと思う」を選択してください。

※インターネット調査のみ

8. 人に気をつかう、やさしい人間だと思う
9. だらしなく、うっかりしていると思う
10. 冷静で、気分が安定していると思う
11. 発想力に欠けた、平凡な人間だと思う

問6 あなたは、平日にスマートフォンやパソコン等を用いて、どの程度の時間インターネットを利用していますか。1日当たりの平均的な利用時間をお答えください。(インターネット利用時間)

インターネットを利用していない方は「0」とお答えください。

1日当たり〇〇時間〇〇分 【1. わからない】

問7◎ 次のインターネット機能・サービスのうち、普段利用しているものをすべてお答えください。(インターネットの利用目的)

1. 電子メールの送受信
2. 天気予報
3. 地図・交通情報の提供サービス
4. ニュースサイト
5. 無料通話アプリやボイスチャット
6. ソーシャルメディア
7. 動画投稿・共有サイト
8. ホームページ・ブログの開設・更新又は閲覧、書き込み
9. 商品・サービスの購入・取引(金融取引及びデジタルコンテンツ購入を除く) 1
0. 辞書・事典サイト
11. オンラインゲーム
12. ラジオ、テレビ番組、映画などのオンデマンド配信サービス

13. デジタルコンテンツの購入・取引
14. クイズ・懸賞応募、アンケート回答
15. 金融取引
16. インターネットオークション
17. e-ラーニング
18. 電子政府・電子自治体
19. その他（電子掲示板やウェブアルバムの利用、電子ファイルの交換など）
20. その他（具体的に：                    ）
21. 利用しない
- 【22. わからない】

問8◎ 現在のお宅の暮らしむきを1年前と比べると、どうでしょうか。この中から1つだけお答えください。（景況感）

1. かなり良い
2. やや良い
3. 変わらない
4. やや悪い
5. かなり悪い
- 【6. わからない】

問9◎ あなたは、現在の生活にどのくらい満足していますか。この中から1つだけお答えください。（生活満足感）

1. 満足している
2. まあ満足している
3. どちらともいえない
4. やや不満だ
5. 不満だ
- 【6. わからない】

問10◎ 多くの人々が「長期的に見ると、自分は〇〇党寄りだ」とお考えのようです。短期的に他の政党へ投票することはもちろんあり得るとして、長い目で見ると、あなたは「何党寄り」と言えるでしょうか。この中から1つだけお答えください。（長期的党派性）

1. 自由民主党
2. 公明党
3. 立憲民主党
4. 日本共産党
5. 日本維新の会
6. 国民民主党
7. 社会民主党
8. れいわ新撰組
9. NHK受信料を支払わない方法を教える党
10. その他の政党（具体的に：                    ）
11. そのような政党はない
12. わからない・答えたくない

問 11 あなたは、菅内閣を支持しますか、支持しませんか。(内閣支持)

1. 支持する
2. 支持しない
3. わからない・答えたくない

問 12◎ 現在、あなたには、配偶者はいらっしゃいますか。届け出の有無に関係なく、この中から1つだけお答えください。(配偶者の有無)

1. 未婚
2. 配偶者あり
3. 死別
4. 離別
5. 答えたくない

問 13◎ 2021年3月25日(木)から2021年3月31日(水)までの1週間にお仕事をしましたか。この中から1つだけお答えください。

1. 主に仕事
2. 家事などのほか仕事
3. 通学のかたわら仕事
4. 普段仕事をしているが、この期間は仕事を休んでいた
5. 普段仕事はしていないが、仕事を探していた
6. 家事のみ
7. 通学のみ
8. その他(高齢者など)
- 【9. わからない・答えたくない】

問 14◎ 問 13で1から4までの回答をした方におうかがいします。あなたのお仕事の地位はどれに当てはまりますか。この中から1つだけお答えください。

1. 正規の職員・従業員
2. 労働者派遣事業所の派遣社員
3. パート・アルバイト・その他(契約社員・嘱託を含む)
4. 会社などの役員
5. 自営業主(雇人のある業主、雇人のない業主)、家庭内職者
6. 家族従業者
- 【7. わからない・答えたくない】

問 15 問 13で1から4までの回答をした方におうかがいします。あなたのお仕事の種類はどれに当てはまりますか。この中から1つだけお答えください。

1. 農業、林業、漁業
2. 鉱業、採石業、砂利採取業、建設業(金属や原油などの採掘、土木や建築全般)
3. 製造業、電気・ガス・熱供給・水道業(食品や工業製品の製造、発電所や上下水道など)

4. 情報通信業（通信や放送、新聞や出版物の制作など）
  5. 運輸業、郵便業（鉄道、道路運送、倉庫、郵便事業など）※郵便局は 14. 複合サービス業へ
  6. 卸売業、小売業（各種卸売や小売、商社や貿易など）
  7. 金融業、保険業（銀行や金融商品取引、各種保険サービスなど）
  8. 不動産業、物品賃貸業（土地・建物の売買や賃貸、各種リースなど）
  9. 学術研究、専門・技術サービス業（研究機関、法律事務所、総合広告サービス、建築設計など）
  10. 宿泊業、飲食サービス業
  11. 生活関連サービス業、娯楽業（洗濯、理美容、冠婚葬祭、旅行業、各種娯楽施設など）
  12. 教育、学習支援業（各種学校、図書館、学習塾、稽古（けいこ）、スポーツ教室など）
- ※保育園は 13. 医療、福祉へ
13. 医療、福祉（病院や診療所、保健衛生施設、保育園、介護施設など）
  14. 複合サービス業、サービス業（他に分類されないもの）  
（郵便局、農林漁協、自動車・機械の整備、警備、宗教団体、講演会や集会の施設運営など）
  15. 公務（他に分類されるものを除く）（行政・司法・立法の非現業公務員全般や警察署など）

※学校、図書館、医療、また、現業の方は 1～14 の中から選んでください

16. その他
- 【17. わからない・答えたくない】

問 16◎ あなたの、お住まいの種類についてお教えてください。この中から 1 つだけお答えください。（住居の種類）

1. 持ち家
2. 公営の借家、都市機構・公社の借家
3. 民営の借家
4. 給与住宅、間借り
5. 会社等の独身寮・寄宿舎
6. その他
- 【7. わからない・答えたくない】

問 17◎ あなたは、5 年前にはどこに住んでいましたか。この中から 1 つだけお答えください。（5 年前の常住地）

1. 現在と同じ住所
2. 同じ区・市町村内の他の場所
3. 他の区・市町村
4. 外国
- 【5. わからない・答えたくない】

## 付録 B 回帰分析における変数の定義

- ・女性ダミー

男性=0、女性=1をとるダミー変数。

- ・年齢

年齢（各歳）を示す変数。

- ・既婚

配偶者の有無について、既婚者=1、それ以外=0を示す変数。

- ・死別

配偶者の有無について、死別=1、それ以外=0を示す変数。

- ・無職

労働力状態について、就業者=0、それ以外を=1とする変数。

- ・家事

労働力状態について、家事=1、それ以外を=0とする変数。

- ・パート

従業上の地位について、雇用者（パート・アルバイト・その他）=1、それ以外=0とする変数。

- ・自営

従業上の地位について、自営業主（雇人のある業主）・自営業主（雇人のない業主）・家庭内職者=1、それ以外=0とする変数。

- ・1次産業

産業大分類について、農業・林業・漁業であれば=1、それ以外=0とする変数。

- ・2次産業

産業大分類について、鉱業、採石業、砂利採取業・建設業・製造業・電気・ガス・熱供給・水道業であれば=1、それ以外=0とする変数。

- ・持ち家

住居の種類・住宅の所有について、持ち家であれば=1、それ以外を=0 とする変数。

・ 常住

5年前の常住地について、現住所であれば=1、それ以外を=0 とする変数。

・ 北海道

居住都道府県について、北海道であれば=1、それ以外を=0 とする変数。

・ 東北

居住都道府県について、青森県・岩手県・宮城県・秋田県・山形県・福島県であれば=1、それ以外を=0 とする変数。

・ 関東

居住都道府県について、茨城県・栃木県・群馬県・埼玉県・千葉県・東京都・神奈川県であれば=1、それ以外を=0 とする変数。

・ 甲信越

居住都道府県について、新潟県・山梨県・長野県であれば=1、それ以外を=0 とする変数。

・ 北陸

居住都道府県について、富山県・石川県・福井県であれば=1、それ以外を=0 とする変数。

・ 東海

居住都道府県について、岐阜県・静岡県・愛知県・三重県であれば=1、それ以外を=0 とする変数。

・ 関西

居住都道府県について、滋賀県・京都府・大阪府・兵庫県・奈良県・和歌山県であれば=1、それ以外を=0 とする変数。

・ 中国

居住都道府県について、鳥取県・島根県・岡山県・広島県・山口県であれば=1、それ以外を=0 とする変数。



・四国

居住都道府県について、徳島県・香川県・愛媛県・高知県であれば=1、それ以外を=0 とする変数。

・九州

居住都道府県について、福岡県・佐賀県・長崎県・熊本県・大分県・宮崎県・鹿児島県・沖縄県であれば=1、それ以外を=0 とする変数。

・21 大都市

21 大都市居住者であれば=1、それ以外を=0 とする変数。

・教育程度

1=（新制）中学／（旧制）小・高小、2=（新制）高校／（旧制）中学、3=（新制）短大・高専／（旧制）高校・専門、4=（新制）大学・大学院／（旧制）大学・大学院 とする変数。

・外向性

・協調性

・誠実性

・神経症傾向

・開放性

各変数 2~14 の値をとり、値が大きいほど各特性が「高い」とする。

・インターネット利用時間

インターネット利用時間（分単位）に対し、対数変換を行った変数。

・インターネット利用目的：ネット1

インターネット利用目的について、「電子メールの送受信」、「天気予報」、「地図・交通情報の提供サービス」、「ニュースサイト」、「商品・サービスの購入・取引（金融取引及びデジタルコンテンツ購入を除く）」、「辞書・事典サイト」のうち、いくつを選択しているかを示す変数。

・インターネット利用目的：ネット2

インターネット利用目的について、「デジタルコンテンツの購入・取引」、「金融取引」、「インターネットオークション」、「電子政府・電子自治体」のうち、いくつを選択しているかを示す変数。

・インターネット利用目的：ネット3

インターネット利用目的について、「ソーシャルメディア」、「動画投稿・共有サイト」、「オンラインゲーム」、「ラジオ」のうち、いくつを選択しているかを示す変数。

・インターネット利用目的：ネット4

インターネット利用目的について、「クイズ・懸賞応募、アンケート回答」を選択していれば=1、それ以外を=0 とする変数。

## 著者

谷口 将紀 NIRA 総合研究開発機構理事長／東京大学教授

大森 翔子 NIRA 総合研究開発機構研究コーディネーター・研究員



---

**NIRA 研究報告書 2021 No.11**  
**インターネット調査におけるバイアス：**  
**国勢調査・面接調査を利用した比較検討**

---

発行 2022年3月  
公益財団法人 NIRA 総合研究開発機構  
〒150-6034 東京都渋谷区恵比寿 4-20-3  
恵比寿ガーデンプレイスタワー34階  
電話 03(5448)1710  
ホームページ <http://www.nira.or.jp/>

---

© NIRA 総合研究開発機構 2022

ISBN: 978-4-7955-0586-5



## NIRA 総研とは

NIRA 総合研究開発機構（略称：NIRA 総研）は、

わが国の経済社会の活性化・発展のために

大胆かつタイムリーに政策課題の論点などを提供する

民間の独立した研究機関です。

学者や研究者、専門家のネットワークを活かして、

公正・中立な立場から公益性の高い活動を行い、

わが国の政策論議をいっそう活性化し、政策形成過程に

貢献していくことを目指しています。

研究分野としては、国内の経済社会政策、国際関係、

地域に関する課題をとりあげます。