

Web 調査結果の信頼性を向上させる方法の検討

片瀬 拓弥

How to Improve the Reliability of Web Survey Results

Takuya KATASE

要旨

本研究は、日本学術会議の Web 調査に対する提言を踏まえ、Web 調査結果の信頼性を向上させる方法を検討した。1つ目は、異なる Web 調査会社を使った全国調査を行い、登録モニターの選択バイアスの影響を検討した。2つ目は、不適切回答者の除去方法として、選択肢の度数分布表やレア体験スケールを活用した。さらに、この除去方法を検証し、Web 調査の信頼性を向上させる方法の留意点をまとめた。

キーワード：Web 調査、選択バイアス、不適切回答者、度数分布表、レア体験スケール

1. はじめに

現代の Web 調査の興隆に対し、日本学術会議（2020）は、「特性を理解せずに安易に利用することは、Web 調査に対する信用を低下させ、有効な学術的利用を阻害しかねない。このことを背景とし、本提言では、Web 調査の利点と問題点を検討した上で、Web 調査の有効な学術的活用のための提言を行う。」としている。この中で「従来型の社会調査を用いてきた社会学者も Web 調査を用いるようになってきていること、従来型社会調査において回収率が低下している」と従来型社会調査の近況を挙げながら、「Web 調査は必要な標本数を確保できるという利点を持っているが、標本の代表性は保証されない」という点を指摘している（日本学術会議 2020）。ただし、「Web 調査の利点を考慮するならば、無作為標本を用いない Web 調査は学術的に意味がないという単純な議論をすることはできない。むしろ Web 調査利用者は本提言で論ずる Web 調査の問題点を的確に理解した上で活用すべきである。」とも述べている。さらに「Web 調査に応じる人と慎重になって応じない人との間で生じる選択バイアスにも注意を払う必要がある。」としている。選択バイアスとは、「本来調査の対象とする集団から一部の対象者が選択されている、あるいは脱落している状況で、単純な解析を行うことで生じる分析結果の歪みのことを指す（星野 2009）」。つまり、Web 調査を活用する際には、「従来型社会調査のような無作為抽出を実施していないこと」から生ずる「選択バイアスの影響」に十分配慮すべきである。

次に日本学術会議（2020）は、「Web 調査において、Satisfice 行動を見せる回答者の存在が指摘される」としている。この Satisfice 行動とは、「satisfy と suffice の 2 単語の組み合わせに由来する造語で、努力の最小化などと訳され、Web 調査の文脈では、質問を含む調査票の内容に十分な注意を払わず、素早く完遂すること（だけ）を目的として回答するような行動を指している」。この Satisfice 行動は、「オンライン調査で測定される態度項目がしばしばマトリックス形式で提示され、回答行動が単調になりがちであることから発生している可能性がある（三浦・小林 2015）」。つまり、Web 調査では、このような不適切回答者が存在することにより、データの信頼性が低下する可能性がある。日本でも不適切回答者を除去する方法が研究されている（三浦・小林 2015, 2018）。三浦・小林（2018）は、海

外研究の成果から IMC と DQS という回答方法を本調査に混在させることで、不適切回答者を除去する方法を報告している。IMC とは Instructional manipulation check の略であり、「設問への回答の際にそれに付随する教示文を精読しない努力の最小化を検出するための方法」であり、また、DQS とは Directed Questions Scale の略であり、心理学で良く用いられるリッカート法などの「尺度項目を精読しない努力の最小化を検出するための方法である（三浦・小林 2018）」。この2つの方法は、不適切回答者を除去するには優れた方法と考えられるが、その一方、適切回答者をも疑うような方法であるため、回答時に不愉快な気持ちを抱かせてしまう可能性は否めない。すなわち、回答者に不愉快な気持ちを抱かせず、不適切回答者を除去する方法の検討も必要であると考えられる。

以上を踏まえ、本研究では、Web 調査を活用するにあたり、その調査結果の信頼性を向上させるため、「登録モニターの選択バイアスの影響」と「不適切回答者の除去方法」に焦点を置くことにした。

2. 本研究の目的

本研究の目的は、日本学術会議（2020）の Web 調査に対する提言を踏まえ、Web 調査結果の信頼性を向上させる方法を検討する。1つ目は、異なる Web 調査会社を使った全国調査を行い、登録モニターの選択バイアスの影響を検討する。2つ目は、不適切回答者の除去方法として、選択肢の度数分布表とリア体験スケール（後述）を活用する。さらに、この除去方法を検証し、Web 調査の信頼性を向上させる方法の留意点をまとめたい。

3. 本研究の方法

（1）調査方法の概要

Web 調査の一般的状況として、総務省の通信利用動向調査（2021）によれば、「インターネット利用者の割合は、13～59歳の各年齢層で9割を超えている。個人のインターネット利用機器は、スマートフォンがパソコンを上回り、20～39歳の各年齢層で9割以上が利用している。」としている。また、60代は82.7%、70代は59.6%となっており、60代までは、インターネット利用者が多数を占めている。つまり、インターネットを利用していることだけを理由にして、「データの代表性に大きな課題がある（選択バイアスの影響がある）」とは言い切れない状況になっている。この点は、日本学術会議（2020）でも指摘している。しかし、登録モニターを用いた Web 調査においては、「登録モニターに選択バイアス」があることは否めない。

本研究では、登録モニターを用いた Web 調査において、「選択バイアス」の影響を軽減させるため、異なる Web 調査会社を通じて、各 5000 名（男女 20 代以上）の全国調査を実施することにした。2つの Web 調査会社を A 社、B 社とした時、調査前に両社の登録モニターの統計資料を提示してもらい、登録モニターとインターネット利用者層に著しい偏りがないことを確認した。さらに調査時には、年代 5 区分（20 代、30 代、40 代、50 代、60 代以上）と性別 2 区分（男女）の計 10 区分を各 500 名とし、全国地域別人口比割で回答を求めるように依頼した。この手順により、登録モニターの偏りを可能な限り軽減するよう努めた。つまり、複数の Web 調査会社を活用することで、少なくとも 1 社のみで調査結果を考察するより、信頼性や再現性が向上するはずである。A 社、B 社の調査は、もともと別の目的（片瀬 2021a、片瀬 2021b）で実施されたものであるが、「不適切回答者を除去するための質問群」は、ほぼ同様の内容として実施した。

また、調査会社からの納品データは、個人情報は一切含まれていない匿名データである。さらに、

研究倫理面を配慮し、回答前に研究同意を取得している。よって、個人情報や人権保護への対応は十分であると考え。両社調査とも、質問群①（目的：選択バイアスの影響を確認、不適切回答者を除去）、質問群②（目的：不適切回答者を除去、不適切回答者の除去方法を検証）、質問群③（目的：不適切回答者の除去）、という3つの質問群を用いる。質問群①及び②は、7件法（全くそう思わない、そう思わない、ややそう思わない、どちらともいえない、ややそう思う、そう思う、強くそう思う）の質問が各10問、質問群③は、2件法（はい、いいえ）の質問が9問で構成された。ただし、質問群③の文言は、A社とB社で一部変更した。

（2）質問群①の調査について（A社、B社共通）

この調査は、A社調査により片瀬（2021a）が作成した対AI信頼感尺度を用いる。また、B社調査の片瀬（2021b）でも用いており、選択バイアスの影響を確認、不適切回答者を除去するための指標として用いる。この質問群は、「AI（人工知能）について、あなたはどのように考えているかお答えください。」という問いに対し7件法で回答し、人のAIに対する信頼感を調査するものである。上5項目を第1因子「AIの社会有益性への信頼感」、下5項目を第2因子「AIの忠実性への不信感（逆転項目）」の2因子構造としている。これらの項目について、ランダム順に出題した。*は逆転項目を示す。

- ・AI（人工知能）は、社会に有益なものである
- ・AI（人工知能）の未来予測は、人の役に立つ
- ・AI（人工知能）は、概ね正しい結果を導き出す
- ・AI（人工知能）は、人の信頼に値するものである
- ・AI（人工知能）は、基本的には正直である
- ・AI（人工知能）は、人を騙（だま）すことがある（*）
- ・AI（人工知能）は、うそをつくことがある（*）
- ・AI（人工知能）は、人の指示に従わないことがある（*）
- ・AI（人工知能）は、人が知らないところで多くの過ちを犯している（*）
- ・AI（人工知能）は、自らの危機回避のために人の指示に従わないこともある（*）

（3）質問群②の調査について（A社、B社共通）

この調査は、小塩ら（2012）が作成した日本語版 Ten Item Personality Inventory（TIPI-J）を用いる。不適切回答者を除去、さらに除去方法を検証するための指標として用いる。TIPI-Jは、Big Fiveの5つの因子を各2項目（逆転項目とペア）で測定する尺度である。Big Fiveとは、外向性、協調性、勤勉性、神経症傾向、開放性の5因子で表される性格検査のことである。この質問群は、「あなた自身の性格について、どのように考えているかお答えください。」の問いに対して7件法で回答する。本研究では、以下のように、小塩ら（2012）と同じ順に出題した。*は逆転項目を示す。

- ・活発で外向的だと思う（外向性）
- ・他人に不満を持ち、もめごとを起こしやすいと思う（*協調性）
- ・しっかりしていて、自分に厳しいと思う（勤勉性）
- ・心配性で、うろたえやすいと思う（神経症傾向）
- ・新しいことが好きで、変わった考えを持つと思う（開放性）
- ・ひかえ目で、おとなしいと思う（*外向性）
- ・人に気をつかう、やさしい人間だと思う（協調性）
- ・だらしくなく、うっかりしていると思う（*勤勉性）

・冷静で、気分が安定していると思う（*神経症傾向）

・発想力に欠けた、平凡な人間だと思う（*開放性）

（4）質問群③（レア体験スケール）の調査について

この調査は、「今までのあなたの人生の中で、以下の体験があったかを教えてください。」の問いに対して、「はい、いいえ」の2件法で回答するものである。不適切回答者を除去するための指標として用いる。以下の内容は、一般的に人生において稀にしか起こらないと考えられるレア体験等である。このような質問に対し、不適切回答者は、虚偽、興味本位、ひやかし等で「はい」と回答をする可能性が高いと考えられる。また、この質問群を「レア体験スケール」と称する。以下の項目について、ランダム順に出題した。

A社 質問項目	B社 質問項目
心肺停止になった経験がある	心肺停止になったことがある
幽霊を見た経験がある	幽霊を見たことがある
幽霊と話した経験がある	幽霊と話したことがある
臨死体験をしたことがある（例えば、三途の川を見た、体外離脱した、先祖のお迎えなど）	臨死体験をしたことがある（例えば、三途の川を見た、体外離脱した、先祖のお迎えなど）
自分の生物学的な性別（男女）に違和感を感じたことがある	自分の生物学的な性別（男女）に違和感を感じたことがある
正夢を見たことがある	UFO（未確認飛行物体）を見たことがある
虫の知らせ（良くないことが起こりそうな予感）を感じたことがある	何らかの中毒症状（薬物、毒物、化学物質等）を起こしたことがある
周りに誰もいない（誰も呼んでいない）のに誰かの声を聞いたことがある	心療内科や精神科に通院したことがある
科学的に理解できない何らかの実体験がある	現在、何らかの宗教を信じている

（5）不適切回答者の除去方法について

度数分布表を活用して不適切回答者を除去する方法について述べる。この方法は、質問群①及び②内において「同じ番号の回答を繰り返し選んでいる者」、「同じ番号の回答を繰り返し選んでいる傾向にあるが、稀に違う回答を選択している者」などを「努力の最小化」を行っている不適切回答者として除去する。

まず、7件法で得られる選択肢1から選択肢7の選択肢ごとの度数分布表を作成する。質問群①及び②を合わせて20個の質問項目があるため、各選択肢の度数は、0から20までとなる。例えば、選択肢1の「全くそう思わない」を20個選択している場合、マトリックス形式の回答群から一番左端か右端の選択肢を連続して選択しているものと考えられ、先に述べた「努力の最小化」が起きていると考える。同様に他選択肢でも同じ選択肢を20個選択していたら、「努力の最小化」が起きていると考える。また、選択肢1の「全くそう思わない」や選択肢7の「強くそう思う」は、質問内容の性質上、確率的に度数が少なくなるはずである。この考え方から、同じ選択肢を一定以上の度数で選択していた回答者を「不適切回答者」とする。そして、この一定以上の基準について、度数分布表の 2σ 以上とすることにした。基準値を 2σ 以上としたのは、外れ値の指標として使われることが多いからである（大久保 2011）。選択肢1から選択肢7までの「不適切回答者」に不適切フラグを立て、少な

表 1 BigFive 尺度因子 (TIPI-J) の項目間相関 (DQS への反応例)

因子	項目	小塩ら (2012)	調査 A		調査 B	
			遵守 (n=5318)	違反 (n=842)	遵守 (n=1173)	違反 (n=58)
外向性	活発で、外向的だと思う	-0.59	-0.44	0.08	-0.53	-0.05
	ひかえめで、おとなしいと思う					
協調性	他人に不満をもち、もめごとを起こしやすいと思う	-0.22	-0.20	0.17	-0.19	0.16
	人に気をつかう、やさしい人間だと思う					
勤勉性	しっかりしていて、自分に厳しいと思う	-0.38	-0.37	0.20	-0.42	-0.14
	だらしなく、うっかりしていると思う					
神経症傾向	心配性で、うろたえやすいと思う	-0.28	-0.30	0.18	-0.31	0.15
	冷静で、気分が安定している					
開放性	新しいことが好きで、変わった考えをもつと思う	-0.39	-0.28	0.17	-0.37	0.19
	発想力に欠けた、平凡な人間だと思う					

※三浦・小林 (2018) の表 3 より転載

くとも 1 つ以上の選択肢で不適切フラグが立った回答者を除去することにした。さらに、質問ごとの除去前と除去後のヒストグラムを見ながら、度数分布に歪みがないかを確認する。一方、質問群③については、「はい、いいえ」の 2 件法であるため、回答者ごと「はい」の回答数を合計し、度数分布表を作成した。この質問群でも「はい」の回答度数が、 2σ 以上の回答者を「不適切回答者」とする。そして、質問群①及び②、質問群③のいずれかで不適切フラグが立った回答者全てを除去することにした。さらに、質問群①及び②、質問群③の両方とも不適切フラグが立った回答者は「データ毀損の意図が明らかな回答者」であり、「特に不適切 (悪質)」と考えられる。

本研究では、このような除去方法を活用し、最終的に全回答者の内、「不適切回答者」とされた者を除いた回答者を「適切回答者」とする。

(6) 不適切回答者の除去方法の検証について

三浦・小林 (2018) は、小塩ら (2012) の TIPI-J を活用し、DQS への指示を遵守した者 (適切回答者と推測される) と遵守しなかった者 (違反者: 不適切回答者と推測される) の性格因子ごとの相関係数を比較した。表 1 に、その結果を転載する。表 1 によれば、遵守群と違反群で比較すると、項目間の相関係数について、違反群の相関係数は、無相関か小塩ら (2012) の相関係数値と差異がある。本研究においても、この TIPI-J を活用し、適切回答者と不適切回答者に分けて項目間相関係数を算出することにより、除去方法の検証を行うことにした。

4. 本研究の結果

(1) 調査結果の概要

本研究では、Web 調査会社の A 社、B 社に対し、年代 5 区分 (20 代、30 代、40 代、50 代、60 代以上) と性別 2 区分 (男女) の計 10 区分を各 500 名とし、全国地域別人口比割で回答を求めるように依頼した。表 2 に、各社納品データを示す。A 社は、依頼した通りの納品数となったが、B 社は、サンプル数が依頼した通りにはならなかった。その理由として、納品前のデータクリーニングにより、サンプル数が依頼した納品数に達しなかったと報告があった。再度、依頼した納品数にするための再調査実施の打診もあったが、納品後にもデータクリーニングを行うため、この納品数で了承した。

表2 A社とB社の納品データ

A社の納品データ				B社の納品データ			
年代	性別		合計	年代	性別		合計
	男	女			男	女	
20代	500	500	1,000	20代	495	497	992
30代	500	500	1,000	30代	484	489	973
40代	500	500	1,000	40代	493	464	957
50代	500	500	1,000	50代	501	498	999
60代以上	500	500	1,000	60代以上	500	500	1,000
合計	2,500	2,500	5,000	合計	2,473	2,448	4,921
	50.0%	50.0%	100.0%		50.3%	49.7%	100.0%

表3 各社納品データから算出した度数分布表の基準値（2σ）と不適切回答者の除去数

調査会社	質問群①と質問群②							質問群③	不適切回答者数		
	選択肢1	選択肢2	選択肢3	選択肢4	選択肢5	選択肢6	選択肢7		総除去数	(悪質)	
A社	基準値	4	6	7	16	10	7	4	5	2,015	(115)
	除去数	275	274	386	316	310	243	244	298		
B社	基準値	5	6	8	15	10	7	4	4	1,901	(115)
	除去数	226	312	291	278	339	282	312	233		

※各除去数は延べ数であり、重複した者も含まれるため、単に合計しても総除去数とは一致しない

表4 A社とB社の適切回答者一覧

A社の適切回答者				B社の適切回答者			
年代	性別		合計	年代	性別		合計
	男	女			男	女	
20代	236	304	540	20代	226	265	491
30代	230	304	534	30代	264	310	574
40代	269	334	603	40代	276	303	579
50代	291	348	639	50代	313	355	668
60代以上	316	353	669	60代以上	338	370	708
合計	1,342	1,643	2,985	合計	1,417	1,603	3,020
	45.0%	55.0%	100.0%		46.9%	53.1%	100.0%

(2) 不適切回答者の除去結果について

第一に、各社納品データを用いて、度数分布表を作成した。そして、その度数分布表から算出した基準値（2σ）と不適切回答者の除去数を表3に示す。表3の総除去数の内、質問群①及び②からの不適切フラグと質問群③の不適切フラグの両方が立った回答者を「特に不適切（悪質）」と表示した。さらに、表4に各社納品データから不適切回答者を除去後の適切回答者数一覧を示す。表4によれば、A社の適切回答者割合は59.7%（2985/5000）、B社の適切回答者割合は61.4%（3020/4921）となり、両社ともに約4割の回答者を「不適切」とした。データ毀損量が約4割となったことは、度数分布表を活用していることから生じていると考えられる。なぜなら、実際には「不適切」とされた回答者の中に「適切回答者」が含まれている可能性が存在するからである。本研究のように「不適切回答者の除去方法」として度数分布表を用いることは、データ毀損量が比較的大きくなるが、その一方で「適切回答者」と推測される群の中に「不適切回答者」が含まれる可能性を低くでき、信頼性が向上すると考えられる。この検証結果については後述する。いずれにしても、分析結果の信頼性向上の観点から、Web調査の生データをそのまま分析することは「かなり危険」であり、分析前にデータクリーニング処理を綿密に行うことが必要である。また、両社データとも年代が低くなるほど「適切回答者」が少なくなることから、年代別回答傾向があると考えられる。Web調査の対象者を選択する際には、この年代別回答傾向にも留意する必要がある。

(3) 登録モニターの選択バイアスの影響について

表5に適切回答者の質問群①（対AI信頼感尺度）に対する統計量比較（A社、B社）を示す。表5によれば、A社、B社の統計量において、3項目のみ平均値の差が有意である。しかし、その効果量は、いずれも0.2未満と小さく、単に標本数が多いことによる有意結果と考えられる。また、各因子のクロンバック α 係数や因子負荷量についても、両社に大きな差があるとは言えず、統計的に妥当な範囲内に収まっていると考えられる。つまり、A社、B社の登録モニターは偏った集団とはいえず、質問群①の尺度において再現性があった。この結果から、適切回答者の各社登録モニターの選択バイアスの影響は、小さいと考えても良いのではなかろうか。

一方、不適切回答者の各社、各因子のクロンバック α 係数や因子負荷量について計算した結果を（クロンバック α 係数、因子負荷量）の形式で表すと、A社：第1因子（0.86、0.68-0.81）、第2因子（0.86、0.67-0.81）、B社：第1因子（0.90、0.77-0.83）、第2因子（0.89、0.71-0.86）となった。各社とも適切回答者の統計量よりも、比較的高い数値となっている。これらは、一見、適切回答者よりも妥

表5 適切回答者の質問群①（対AI信頼感尺度）に対する統計量比較（A社、B社）

質問項目		A社	B社	平均値差	効果量
		n=2985	n=3020	t値	Cohenのd
第1因子:AIの社会有益性への信頼感	クロンバック α	0.72	0.76	-	-
	因子負荷量の範囲	.47-.73	.53-.71		
AI（人工知能）は、社会に有益なものである	M (σ)	4.75 (0.95)	4.74 (0.92)	.40	.01
AI（人工知能）の未来予測は、人の役に立つ	M (σ)	4.79 (0.94)	4.77 (0.95)	.81	.02
AI（人工知能）は、概ね正しい結果を導き出す	M (σ)	4.35 (0.90)	4.37 (0.92)	.71	.02
AI（人工知能）は、人の信頼に値するものである	M (σ)	4.17 (0.85)	4.23 (0.91)	2.41*	.06
AI（人工知能）は、基本的には正直である	M (σ)	4.61 (0.97)	4.62 (0.95)	.42	.01
第2因子:AIの忠実性への不信感	クロンバック α	0.67	0.74	-	-
	因子負荷量の範囲	.37-.61	.45-.77		
AI（人工知能）は、人を騙（だま）すことがある	M (σ)	3.92 (0.95)	3.84 (1.00)	2.99**	.08
AI（人工知能）は、うそをつくことがある	M (σ)	3.80 (0.97)	3.83 (0.99)	1.43	.04
AI（人工知能）は、人の指示に従わないことがある	M (σ)	4.20 (0.94)	4.17 (0.93)	1.26	.03
AI（人工知能）は、人が知らないところで多くの過ちを犯している	M (σ)	3.77 (0.95)	3.91 (0.89)	5.96***	.15
AI（人工知能）は、自らの危機回避のために人の指示に従わないこともある	M (σ)	4.06 (0.97)	4.06 (0.94)	.04	.00

* $p < .05$, ** $p < .01$, *** $p < .001$

表6 BigFive 尺度因子（TIPI-J）の項目間相関による不適切回答者の除去方法の検証

因子	項目	小塩ら (2012)	A社			B社		
			適切 n=2985	不適切		適切 n=3020	不適切	
			全体 n=2015	悪質 n=115	全体 n=1901	悪質 n=115		
外向性	活発で外向的だと思う	-0.59	-0.45	-0.23	0.00	-0.48	-0.20	0.23
	ひかえ目で、おとなしいと思う							
協調性	他人に不満を持ち、もめごとを起こしやすいと思う	-0.22	-0.22	-0.05	0.11	-0.30	-0.15	0.26
	人に気をつかう、やさしい人間だと思う							
勤勉性	しっかりしていて、自分に厳しいと思う	-0.38	-0.36	-0.18	0.12	-0.46	-0.24	0.47
	だらしない、うっかりしていると思う							
神経症傾向	心配性で、うろたえやすいと思う	-0.28	-0.34	-0.21	0.00	-0.44	-0.19	0.38
	冷静で、気分が安定していると思う							
開放性	新しいことが好きで、変わった考えを持つと思う	-0.39	-0.30	-0.03	0.19	-0.34	-0.10	0.54
	発想力に欠けた、平凡な人間だと思う							

当性が高いようにみえてしまう。しかし、この結果は「不適切回答者が同じ選択肢を何度も選択すること」により起こる疑似的な統計量の高さであり、単にクロンバック α 係数や因子負荷量が高いから、分析結果や各因子が妥当だと考えることの危険性を示唆している。むしろ、あまりに高い「クロンバック α 係数」や「因子負荷量」は、Web 調査のデータクリーニングの観点から不適切回答が含まれている可能性を検討した方が良いかもしれない。

（4）不適切回答者の除去方法の検証について

表6に BigFive 尺度因子 (TIPI-J) の項目間相関による不適切回答者の除去方法の検証結果を示す。表1と、ほぼ同じ表示方法で結果を示した。A社、B社とも適切回答者の項目間相関は、小塩ら(2012)の傾向とほぼ一致しており、不適切回答者をほぼ除去できていることが検証された。一方、不適切回答者の項目間相関は、無相関(灰色部:0.2未満)が多い。特に不適切回答者の中でも悪質回答者は、相関が正となっており、逆転項目の内容を読まずに回答していることが明らかである。ただし、表1と表5を比較すると、本研究の悪質回答者群と三浦・小林(2018)の違反群が対応しているようにみえる。つまり、本研究の不適切回答者全体の項目間相関には、弱い相関(A社:外向性・神経症傾向、B社:外向性・勤勉性)が残っていたことから判断すると、その中に適切回答者も含まれていた可能性がある。本研究では、不適切回答者を推測する条件として、度数分布表の 2σ 以上と厳しく見積もったが、この条件を緩和しながら、 2.5σ や 3σ 以上なども視野に入れて、除去方法の最適化を行うべきかもしれない。

（5）レア体験スケールの統計結果について

表7にレア体験スケールの統計結果(「はい」の回答率)を示す。両社の質問内容が、一部で異なっているため、完全に比較できない状態であるが、共通項目で見ると適切回答者と不適切回答者で統計値にズレがあることが分かる。ただし、適切回答者の回答が10%以上の項目では、不適切回答者とあまり変わりのないものも存在する。このことから、不適切回答者を除去するためには、適切回答者が10%未満と回答した項目内容を残していくべきではないかと考える。さらに、適切回答者が回答した統計値についての妥当性も検討する必要がある。例えば、他文献に掲載されている統計値と比較するという作業である。本研究は、統計値自体の妥当性を検討するという目的ではないため、この作業は割愛するが、レア体験の出現率として興味深い項目もあり、いずれ追及していくつもりである。付録に追加統計資料を掲載した。

表7 レア体験スケールの統計結果(「はい」の回答率)

レア体験スケールの質問項目	A社		B社	
	適切 n=2985	不適切 n=2015	適切 n=3020	不適切 n=1901
心肺停止になったこと(経験)がある	1.0%	4.7%	0.6%	6.8%
幽霊を見たこと(経験)がある	7.7%	19.0%	9.1%	14.9%
幽霊と話したこと(経験)がある	1.1%	6.9%	1.2%	7.4%
臨死体験をしたことがある(例えば、三途の川を見た、体外離脱した、先祖のお迎えなど)	2.2%	8.6%	2.5%	10.6%
自分の生物学的な性別(男女)に違和感を感じたことがある	4.2%	9.6%	3.9%	10.2%
正夢を見たことがある	32.0%	38.3%		
虫の知らせ(良くないことが起こりそうな予感)を感じたことがある	36.5%	38.6%		
周りに誰もいない(誰も呼んでいない)のに誰かの声を聞いたことがある	12.9%	22.1%		
科学的に理解できない何らかの実体験がある	22.9%	28.2%		
UFO(未確認飛行物体)を見たことがある			7.5%	13.9%
何らかの中毒症状(薬物、毒物、化学物質等)を起こしたことがある			2.6%	9.0%
心療内科や精神科に通院したことがある			17.9%	22.6%
現在、何らかの宗教を信じている			7.3%	12.8%

5. まとめ

本研究は、日本学術会議（2020）の Web 調査に対する提言を踏まえ、Web 調査結果の信頼性を向上させる方法を検討した。1つ目は、異なる Web 調査会社を使った全国調査を行い、登録モニターを選択バイアスの影響を検討した。2つ目は、不適切回答者の除去方法として、選択肢の度数分布表やレア体験スケールを活用した。さらに、不適切回答者の除去方法を検証した。

まとめとして、Web 調査の信頼性を向上させる方法の留意点を、①質問紙作成時、②Web 調査依頼時、③データクリーニング処理時、④データ分析時の4つに分け、箇条書きにした。

①質問紙作成時の留意点

- ・IMC や DQS などの不適切回答者を見分ける質問を含めるか検討する
- ・本調査は、逆転項目を含む質問群を設定する（努力の最小化を見分けるため）
- ・本調査とは別に、データ信頼性について検証を行う指標を入れる（TIPI-J やレア体験スケール等）
- ・「努力の最小化」とは反対の「努力の最大化」が必要な記述式質問を入れるか検討する

②Web 調査依頼時の留意点

- ・調査対象者の年代別回答傾向を確認し、必要ならば調査会社と相談する
- ・研究倫理面に配慮した同意確認手段の確保を依頼する
- ・可能な限り、質問群はランダム出題するように依頼する
- ・Web 調査会社に、登録モニターの統計資料を開示してもらう（選択バイアスの影響を確認）
- ・Web 調査会社に、「努力の最小化」傾向のある回答は、一次加工として除去を依頼する
- ・資金的に可能ならば、同じ内容の質問群を他 Web 調査会社にも依頼する（再現性確保）
- ・記述式質問の文字数が多い順に回答を選択してもらう（依頼数以上の回答者があった場合）

③データクリーニング処理時の留意点

- ・IMC や DQS などにより、不適切回答者を除去する
- ・「努力の最小化」傾向のある不適切回答者について、度数分布表等を活用して除去する（必要に応じて、度数分布表の基準値を模索する）

④データ分析時の留意点

- ・データ信頼性について、適切な指標を用いて検証を行う
- ・質問項目ごとにヒストグラムを作成し、分布に不適切な歪みがないことを確認する
- ・記述統計量を計算し、平均値と標準偏差から天井効果やフロア効果がないことを確認する
- ・因子分析の場合、クロンバック α 係数や因子負荷量が高すぎないか（低すぎないか）確認する

以上、Web 調査の信頼性を向上させる方法の留意点をまとめた。ただし、これらの留意点は、本研究に類似した Web 調査に関してのものである。一般的には、これで十分だとは言いきれないため、今後も Web 調査の信頼性を向上させる様々な研究が必要となろう。

謝辞

本研究は、JSPS 科研費 JP18K02882 の助成を受けた。

引用・参考文献

日本学術会議（2020），Web 調査の有効な学術的活用を目指して．社会学委員会（Web 調査の課題に関する検討分科会），<https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t292-3.pdf>（2020.7.10 公表）

星野崇広（2009），調査観察データの統計科学 因果推論・選択バイアス・データ融合．岩波書店
通信利用動向調査（2021），総務省令和2年調査（2021.6.18公表）

<https://www.soumu.go.jp/johotsusintokei/statistics/statistics05a.html>（閲覧日 2021.12.28）

三浦麻子・小林哲郎（2015），オンライン調査モニタの Satisfice 行動に関する実験的研究．社会心理学研究，31: 1-12.

三浦麻子・小林哲郎（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): 1-11.

片瀬拓弥（2021a），人工知能(AI)に対する信頼感尺度の作成と信頼性・妥当性の検討—全国 Web 調査による性別・年代別のコホート分析—. 日本教育工学会研究報告集，2021(3): 172-179.

片瀬拓弥（2021b），心理的 AI デバイドをアセスメントするテストバッテリーの作成．日本教育工学会研究報告集，2021(4): 67-72.

大久保街亜（2011），反応時間分析における外れ値の処理．専修人間科学論集（心理学篇），1: 81-89.

SUMMARY

This study examined methods to improve the reliability of Web survey results, based on the recommendations of the Science Council of Japan on web surveys. First, I conducted a nationwide survey using different Web survey companies to examine the effects of selection bias of registered monitors. The second was to use a frequency distribution table of choices and a rare experience scale as a method of removing inappropriate respondents. Furthermore, I verified this removal method and summarized the points to keep in mind on how to improve the reliability of Web surveys.

Keywords: Web survey, selection bias, inappropriate respondents, frequency distribution table, rare experience scale

付録

付表 A：A 社と B 社の適切回答者を合算し、レア体験スケールの心肺停止と臨死体験の有無に関するクロス表

付表 B：A 社と B 社の適切回答者を合算し、レア体験スケールの性別違和感の有無の年代別クロス表

レア体験の有無	臨死体験		合計	構成比
	なし	あり		
なし	5,827	131	5,958	99.2%
心肺停止	97.8%	2.2%	100.0%	
あり	36	11	47	0.8%
	76.6%	23.4%	100.0%	
合計	5,863	142	6,005	100.0%
	97.6%	2.4%	100.0%	

年代	性別違和感		合計
	なし	あり	
20代	955	76	1,031
	92.6%	7.4%	100.0%
30代	1,057	51	1,108
	95.4%	4.6%	100.0%
40代	1,133	49	1,182
	95.9%	4.1%	100.0%
50代	1,278	29	1,307
	97.8%	2.2%	100.0%
60代以上	1,338	39	1,377
	97.2%	2.8%	100.0%
合計	5,761	244	6,005
	95.9%	4.1%	100.0%