



データとうまく付き合っていくためには

財務総合政策研究所 総務研究部 財政経済計量分析室 研究官
升井 翼

今月のPRI Open Campusでは、本年5月に財務総合政策研究所（以下、「財務総研」）から刊行されたリサーチ・ペーパー（RP）「法人企業景気予測調査の調査結果の継続性の検証」について、どのような問題意識に基づいており、何を明らかにしようとしているのかを「ファイナンス」の読者の方々に紹介します。ポイントは、統計をはじめとする「データ」が、どこまでその調査方法によって影響を受けるのか、ということであり、調査方法が回答に及ぼす実際の事例を交えつつ紹介します。

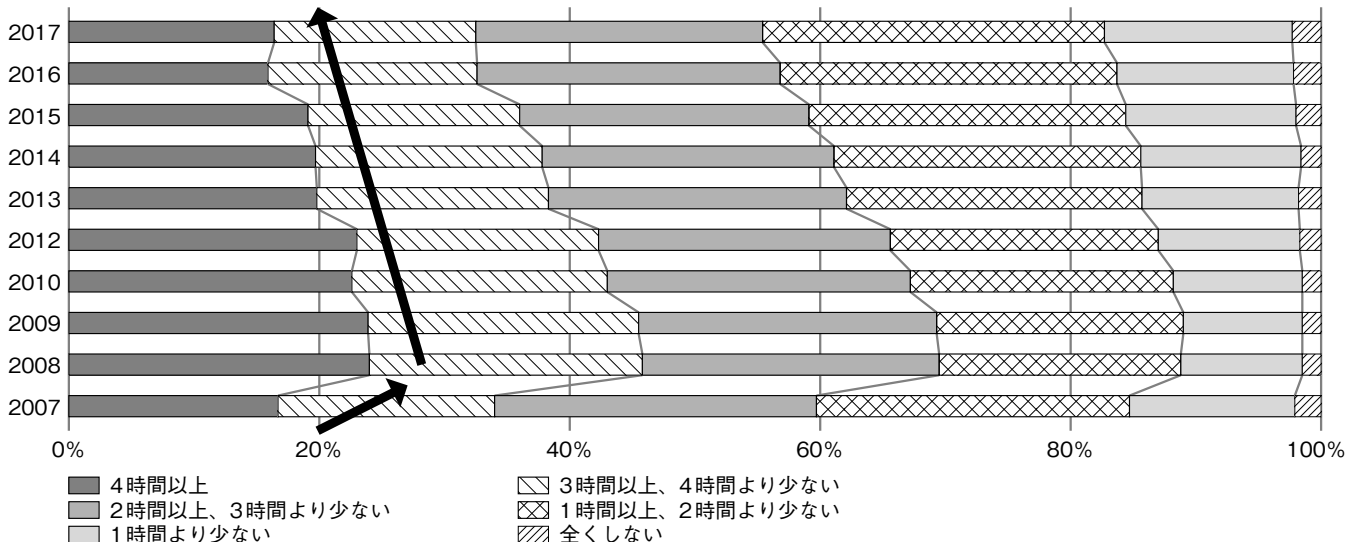
時間より少ない」の割合が増加し、右の「1時間以上、2時間より少ない」と「1時間より少ない」の割合が減少しています。しかし、2008～2017年度の推移を見ると、「4時間以上」と「3時間以上、4時間より少ない」の割合はこの10年で減少傾向にあり、「1時間以上、2時間より少ない」と「1時間より少ない」の割合では逆に増加傾向が読み取れます。2007年度から2008年度にかけての動きが、その後の推移から大きく外れているのです。2008年度以降の動きは、SNS等の普及によるメディアの多様化等によって生活スタイルに占めるテレビ視聴のウェイトが小さくなっていることから説明がつきそうですが、2007年度から2008年度の動きはなぜでしょうか。小学6年生の興味関心があるテレビ番組などが多かったのでしょうか。それとも、児童の生活様式が急に変わったのでしょうか。このころの時代背景は…など、様々なことが頭に浮かびますが、決め手はなさそうです。

1. データ（数字）だけ見ていてはわからないことがある

図表1は小学6年生のテレビ等の視聴時間を調査した結果です。2007～2008年度の推移を見ると、100%積み上げグラフ左の「4時間以上」と「3時間以上、4時間より少ない」と「2時間以上、3時間より少ない」と「1時間以上、2時間より少ない」と「1時間より少ない」の割合が増加し、右の「1時間以上、2時間より少ない」と「1時間より少ない」の割合が減少しています。

図表1 小学6年生のテレビ等の視聴時間

普段（月～金曜日）、1日当たりどれくらいの時間、テレビやビデオ・DVDを見たり、聞いたりしますか
 （勉強のためのテレビやビデオ・DVDを見る時間、テレビゲームをする時間は除く）



（出所）全国学力・学習状況調査から筆者作成

連載
PRI Open
Campus

図表2 テレビ等の視聴時間の回答選択肢
2007年度

- (17) ふだん(月曜日から金曜日)、1日あたりどれくらいの時間、テレビやビデオ・DVDを見たり、聞いたりしますか。(テレビゲームをする時間は除きます。)
- 1 全く見たり、聞いたりしない
 - 2 1時間より少ない
 - 3 1時間以上、2時間より少ない
 - 4 2時間以上、3時間より少ない
 - 5 3時間以上、4時間より少ない
 - 6 4時間以上

(出所) 全国学力・学習状況調査 (2007、2008)

実は、こうした結果が得られたことについては、「いったいどんな面白い番組が流行っていたのかと訝られたが、原因はおそらく調査票の設計にある(土屋2014)」というのが種明かしです。2007年度の調査票と2008年度以降の調査票を見比べてみると明らかな違いがあります(図表2)。

選択肢の並び順が異なっているというのが実は大きな違いです。2007年度の調査票では、1番に「全く見たり、聞いたりしない」が置かれ、そのあと視聴時間が短い順に選択肢が並んでいるのに対し、2008年度以降では選択肢の並び順が真逆になっています。ひとつひとつの回答カテゴリは全く一緒ですので、並び順がどうであれ「正しい」選択肢が選ばれば結果に影響はないはずですが、しかし、選択肢回答方式では、最初に置かれた選択肢ほど選ばれやすい傾向があることは調査の分野では良く知られています(Krosnick & Alwin 1987)。今回の調査に当てはめて考えると、2007年度は比較的短い時間の選択肢が選ばれやすい状況にあり、2008年度以降は比較的長い時間の選択肢が選ばれやすい状況にあったのです。そしてその傾向のおおりの結果が得られているというわけです。

上述の例は回答選択肢の順序が回答に影響を与えているものですが、他にも質問文から受ける印象(林1970)や複数ある質問の回答順序(Schuman & Presser 1981、Herek & Capitanio 1999)、複数回答方式か強制選択方式か(Smyth, Dillman, Christian & Stern 2006)、また調査票のレイアウトによる影響(Christian & Dillman 2004)など様々な要因が議論されています。

2008年度～

- (12) ふだん(月曜日から金曜日)、1日あたりどれくらいの時間、テレビやビデオ・DVDを見たり、聞いたりしますか。(テレビゲームをする時間は除きます。)
- 1 4時間以上
 - 2 3時間以上、4時間より少ない
 - 3 2時間以上、3時間より少ない
 - 4 1時間以上、2時間より少ない
 - 5 1時間より少ない
 - 6 全く見たり、聞いたりしない

こうした調査方法による回答への影響については回答データを見るだけでは気付きにくく、上述の例のように継続した調査において調査方法を変更した場合や予備調査等による検証を経ない限りはインパクトの軽重や本当に影響が生じているのかどうかを含めて判断が困難です。しかし、データに向き合う際には、そのデータがどのように計測されたもので、どのような計測上の傾向が生じうるのかについて考慮する必要があるのです。

2. 「合計」を聞くか「分割」を聞くか

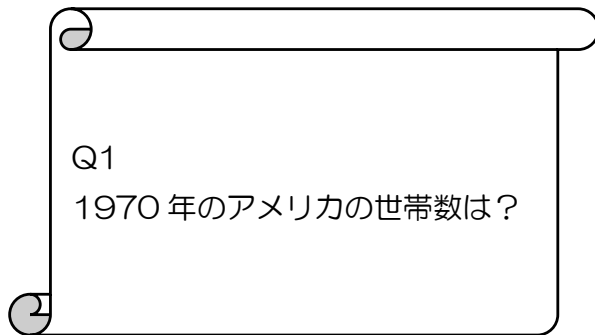
本年5月に財務総研から刊行されたRP「法人企業景気予測調査の調査結果の継続性の検証」は前章の問題意識に関連した研究です。このRPでは質問方法として「まとめて尋ねる(合計)」か「個別に尋ねる(分割)」か、という質問の仕方の違いによって回答に影響が生じるのかを検証したものと位置づけることができます。まずは2.1、2.2で関連する文脈での先行研究を、続いて2.3でRPの概要をご紹介します。

2.1 先行研究(1) 段階的な質問による数量の推定(Armstrong et al. 1975)

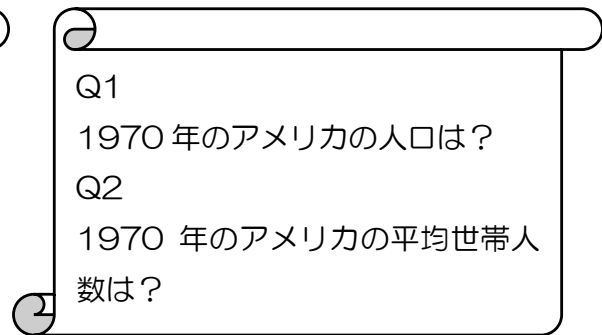
この研究では、クイズのように正答がある問いについて、回答として数量を推定してもらって調査を行っています。その際に図表3のように2通りの質問票を用いて、質問票Aでは世帯数を直接質問しており(Direct version)、質問票Bでは世帯数を算出するために分割された2問を質問しています(Decomposed version)*1。

*1) 回答した(人口÷平均世帯人数)で世帯数を計算する。

図表3 1970年代のアメリカの世帯数を尋ねる2種類の質問票
質問票A



質問票B



(出所) Armstrong, J. S., Denniston Jr, W. B., & Gordon, M. M. (1975) . The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance*, 14 (2) , P257-263. から筆者作成

この研究では1970年代のアメリカについての質問がされていますが、ぜひ日本や身近なことに置き換えてその違いを考えてみてください。

質問票Aより質問票Bの方では問題が小分けになっていて考えやすかったのではないのでしょうか。実際に2種類の調査票を用いた実験の結果も、質問を分割した調査票Bを使用した群の方が正解により近い回答が得られたと報告されています。この結果については、下記の他の質問内容、分割内容、分割数、質問の順序のバリエーションを加えて検証されており、いずれの場合においても細かく分割して聞く方が段階的に推定できることなどにより正確な値が推定できているようです(図表4)。

- ・1970年のアメリカの高校中退者数について(Dropouts)
- ・1970年のアメリカのポラロイドカラーフィルムの消費数について (Film)
- ・1969年のアメリカのタバコ生産量について(Tobacco)
- ・1972年にフィラデルフィアで開催されたコンテストへの応募数について (Contest)

2.2 先行研究(2) 丁寧な想起(Sudman & Bradburn 1973)

もう1つの研究では、1週間のうちに見かけた雑誌について、「どの雑誌を見かけましたか？」と直接質問した場合(Unaided Recall)と、「これらの雑誌を見かけましたか？」と雑誌のリストを見せて質問した

図表4 2種類の質問票による正確性の比較

TABLE 1
Median Accuracy Ratio: Decomposed vs. Direct Questions

Question	(1) Original rank	(2) Direct version (n = 25)	(3) Decomposed version (n = 25)	Difference
Group 1				
Dropouts	2	5.3	2.2	3.1
Film	3	21.1	2.1	19.0
Tobacco	4	103.2	2.1	101.1
Group 2				
Families	1	1.3	1.3	0.0
Dropouts	2	5.4	2.3	3.1
Film	3	10.2	7.0	3.2
Tobacco	4	18.0	7.4	10.6
Contest	5	831.3	147.8	683.5
Group 3				
Families	1	1.4	1.2	0.2
Dropouts	2	5.9	6.0	0.1
Film	3	5.1	2.7	2.4
Tobacco	4	9.0	5.7	3.3
Contest	5	351.0	184.7	166.3

The rankings of differences for Groups I and II were in perfect agreement with the hypothesis (see columns 1 and 4 in the table). While the number of questions for Group I was too small to allow for a test of statistical significance, the rankings in Groups II and III were each significant at the .05 level (using Spearman Rank Correlation from Siegel, 1956). Thus, the results also support H2--i.e., that the value of decomposition is greatest where the subject's knowledge is poorest.

(出所) Armstrong, J. S., Denniston Jr, W. B., & Gordon, M. M. (1975) . The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance*, 14 (2) , P257-263.

場合 (Aided Recall) を比較しています。

読者のみなさんにご想像いただきたいのですが、まず前者の質問方法ではどうでしょうか。何の補助もなく漠然と過去の行動等を思い出すことはそれなりに難易度が高いのではないかと思います。対して後者の質問方法のように雑誌リストが提示されていると、想起がしやすくなっているはずで

す。実験の結果でも、リストを見せて質問をした方が回答される雑誌数が多かったと報告しています(図表5)。記憶に対する調査に対しては、Omission error*²及びTelescoping error*³と呼ばれるエラーが作用すると考えられており、リスト形式で一つずつ判断させるような丁寧な想起を促すと、Omission errorが低減する一方でTelescoping errorが増幅され、総合して回答レベルが増加すると考えられています。

図表5 記録の有無、設問の構成、想起の補助による回答への効果

Item	Average Response Effects		
	Net Mean	σ	N
Records Available:			
Yes	-.03	.28	896
Uncertain	.03	.24	333
No	.06	.78	939
Structure of Questions:			
Closed-ended	-.001	.42	1,750
Open-ended	.01	.64	1,376
Recall:			
Aided	.13	1.31	208
Unaided	-.02	.48	1,209

(出所) Sudman, S., & Bradburn, N. M. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 68 (344), P805-815.

2.3 RP「法人企業景気予測調査の調査結果の継続性の検証」

このRPでは、2つの先行研究等を踏まえ、法人企業景気予測調査の調査方法の変更の影響について検証を実施しています。具体的な変更点は、図表6のように2019年度から企業収益等に関する計数項目の記入単位を前期、後期の2項目から年度の1項目に変更したものです。

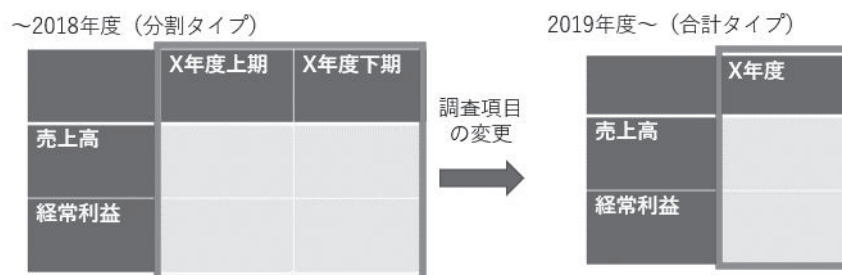
年度の計数を記入する場合と半期ごとの計数を記入する場合には、半期ごとの予測を段階的に考えること、あるいは丁寧に予測することにより年度の計数をまとめて聞くのとは異なるのではないかという疑問が浮かんできます。この点について実際には事前に有識者等へ諮ったうえで調査項目変更を実施しており、影響は軽微であろうことは予想されていましたが、変更してから実際に得られたデータを用いて検証を試みました。

検証方法としては、法人企業景気予測調査と母集団名簿が共通している法人企業統計調査(年次別調査)を利用し、法人企業景気予測調査の調査項目の記入単位の変更前後で両統計による結果*⁴を比較し、変化が生じているか確認しました。

図表7から、平均値及び分散では2018年度以前、2019年度以降ともほぼ一定の間隔を保っています。企業ごとの両調査の計数についての相関係数*⁵については、全期にわたって一貫して非常に高い値となっています。この結果から、調査方法の変更前後の調査期による違いは見られませんでした。

つまり、法人企業景気予測調査においては、調査方

図表6 法人企業景気予測調査の調査項目の変更点



(出所) 法人企業景気予測調査から筆者作成

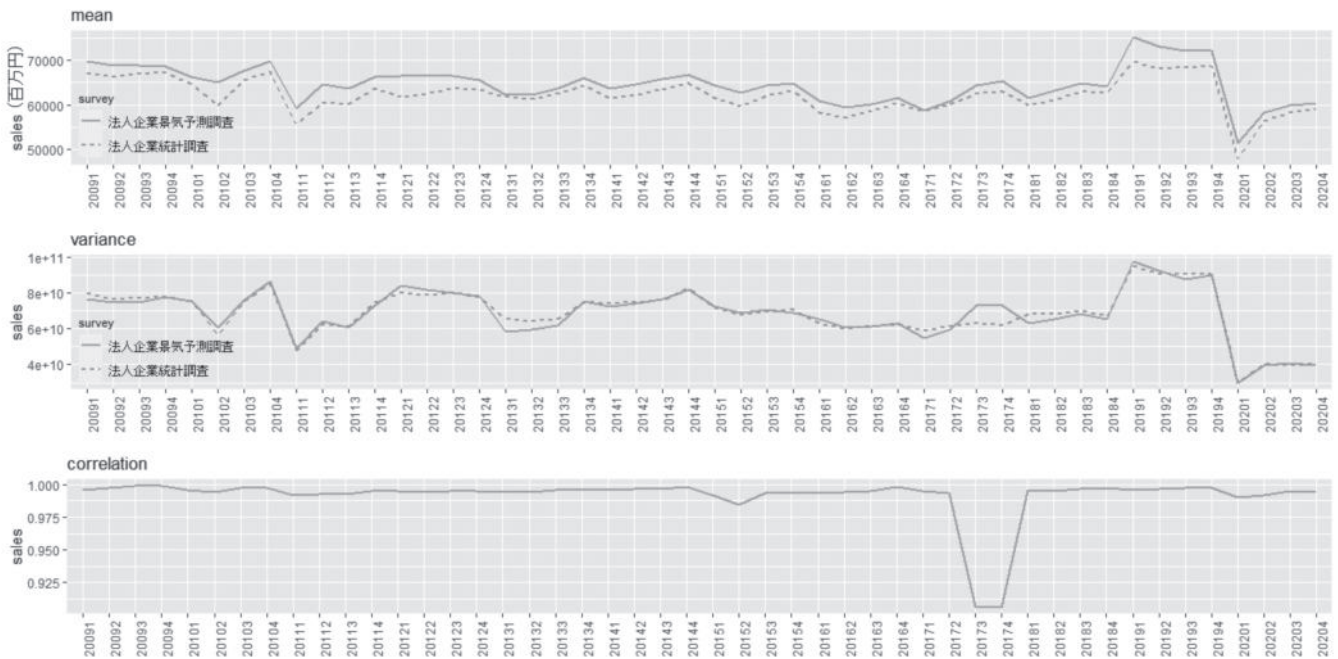
*2) 報告すべきイベント等の数え漏れにより回答レベルが減少する効果。

*3) 報告すべきでないイベント(例えば、直近1週間の出来事の報告を求められたときの、2週間以上前のイベント)を錯覚して報告することにより回答レベルが増加する効果。

*4) 法人企業景気予測調査では売上高等の当期の見通し、法人企業統計調査では売上高等の決算計数を利用した。

*5) 両調査の計数の関連の強さを示し、値が1に近いほど関連が強く、連動していることを示す。法人企業景気予測調査では売上高等の見通しを回答しているため、法人企業統計の決算計数と原理的に一致するものではない。その意味では、各社の両計数についての相対的な位置関係を表す相関にて判断する方が変化をとらえるためにより適していると考えられる。

図表7 法人企業景気予測調査と法人企業統計調査の売上高の推移



(出所) 法人企業景気予測調査、法人企業統計調査から筆者作成

法が変わったものの、そのことによって結果に影響を与えたとはいえないというものでした。つまり調査結果の時系列の推移をみるときに特別な操作なく統計の継続性が保たれており、そのまま分析に用いることができることが確認できたといえます。今回の結果は、素人目には単なる杞憂、取り越し苦労ともみえますが、この研究の貢献として、異なる調査方法が採られたときに回答に変化が生じるケースばかりではなく、変化が生じないケースを示すことができた点があります。ここから、どのような場合に両者が区別されるのか考えていきます。

3. Satisficing～調査方法によって回答が変化する回答者心理

それでは、回答に変化が生じるケースと生じないケースの差異は何でしょうか。

調査方法によって回答が変化するということは、過去の同一の行動等について質問する場合でも、質問形式Aで聞いた場合と質問形式Bで聞いた場合の回答が異なる人が少なからず存在するという事です。なぜ

そのようなことが起こるか、という疑問に対する一つの答えとして、「Satisficing」と呼ばれる回答者心理が考えられています。

Satisficingとはsatisfy（満足させる）とsuffice（十分である）の合成語であり、Simon（1957）によって心理学等の文脈で「目的を達成するために十分な行為」を指す用語として使われました。この考え方がKrosnick & Alwin（1987）、Krosnick（1991）によって調査の文脈に援用され、「満足のいく回答を作成するために最小限の労力を費やすこと」と解釈されています。なお、自計記入式質問紙についての回答の作成プロセスは「認知・理解・検索・判断・表現^{*6}」の5ステップとされており（Dillman et al. 2014）、このステップを完遂させるのではなく一部または全部を省略することがすなわちSatisficing行動だと考えられます。例えば、全て同じ選択肢を選ぶ（Straight lining）、前に示された選択肢を選ぶ（Primacy effect）などの回答行動が知られています。このようなSatisficing行動が採られると調査方法の変更等により回答が変化することがわかります。

Satisficingの起こりやすさにはTask difficulty、Ability及びMotivation^{*7}が関係していて、下記の概

*6) 認知：どの質問に回答すればよいのか認識する。理解：何が問われているのか理解する。検索：記憶や記録等から適切な情報を引き出す。判断：問われていること、引き出した情報を照らし合わせて回答を作成する。表現：要求されている形で回答を表現する。

*7) Task difficulty：質問文や選択肢の解釈、記憶の検索、比較や尺度判断などの困難度。Ability：認知的レベル、特定のテーマ等への造詣などの回答者の能力。Motivation：回答者にとっての質問のテーマの重要度、調査の社会的価値、責任感などの調査に対する熱心さ。

念的な式に表されるように、質問のTask difficultyが高いほど、あるいは回答者のAbility及びMotivationが低いほどSatisficingが起りやすいとされています (Krosnick 1991)。

$$p(\text{Satisficing}) = \frac{a_1(\text{Task difficulty})}{a_2(\text{Ability}) \times a_3(\text{Motivation})}$$

法人企業景気予測調査では、政府統計であることや調査対象が企業経営者であることなどから先に示したAbilityやMotivationが高く、Satisficingが起りづらい特徴があると考えられます。一方、冒頭の全国学力・学習状況調査では、小学6年生を対象にしていることなどからAbilityやMotivationが比較的低かったのかもしれませんが。近年多くの調査で採用されているWeb調査等でも、回答者がポイントなどの報酬を得ることを主目的に日常的に数多くの調査に回答する傾向があることなどからMotivationが低くSatisficingが起りやすい環境であることが考えられます (三浦・小林 2015)。

4. おわりに

本稿では、調査方法のささいに思えるような違いで異なる結果が得られることがあるということ、そして、政府統計やWeb調査等の性質の違いによって影響の受けやすさに違いがあることを具体的な事例を通して紹介してきました。

これを踏まえて、データとうまく付き合っていくためには、調査実施者の立場からは、実際に調査を実施することになったときに、調査方法をどのように設計すればよいのか、知ろうとしているものがその方法で本当に得られているのか、という観点が重要であることがわかります。またデータユーザー (分析者) の立場からは、調査方法等によって結果に影響を生じうるということを念頭に置いたうえで、調査方法が結果に影響を与えていないかに気を配りながら結果を解釈することが求められます。結果の数値のみに着目し、データの内容や調査方法を吟味することなくいきなり

複雑な統計手法等に頼りすぎることは誤りを犯す恐れがあり、そのデータがどのように収集されたものであるか、把握しておくことも重要なのです。

参考文献

- 土屋隆裕 (2014). 事例に見る調査票の設計と回答者の回答行動. *マーケティング・リサーチ*, 125, 24-32.
- 林知己夫 (1970). 身近な社会. 統計数理研究所国民性調査委員会 (編)『第2日本人の国民性』至誠堂, 75-110.
- 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 (1), 1-12.
- Armstrong, J. S., Denniston Jr, W. B., & Gordon, M. M. (1975). The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance*, 14 (2), 257-263.
- Christian, L. M., & Dillman, D. A. (2004). The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions. *Public Opinion Quarterly*, 68 (1), 58-81.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New Jersey: John Wiley & Sons.
- Herek, G. M., & Capitanio, J. P. (1999). Sex differences in how heterosexuals think about lesbians and gay men: Evidence from survey context effects. *Journal of Sex Research*, 36 (4), 348-360.
- Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51 (2), 201-219.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5 (3), 213-236.
- Schuman, H., & Presser, S. (1981). *Questions & Answers in Attitude Surveys*. Massachusetts: Academic Press.
- Simon, H. A. (1957). *Models of Man; Social and Rational*. New Jersey: John Wiley & Sons.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, 70 (1), 66-77.
- Sudman, S., & Bradburn, N. M. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 68 (344), 805-815.

プロフィール

研究官 升井 翼

2013年に北陸財務局に入局。2016年から財務総合政策研究所へ出向し、統計の作成、財政の長期推計などに従事しています。また、在職中の2023年に修士号 (データサイエンス) を取得しました。