

第 2 章

順序尺度を用いた文化の計測：問題点と対処方法の紹介

明日山 陽子

要約：

本章では、文化研究でよく用いられる順序尺度(ordinal scale)を使用して個人や集団間の平均値を比較したり、回帰分析を行ったりする際に、どのような問題が生じるのか、またそれらの問題にどう対処すればよいのか、既存研究をもとに検討する。順序尺度の番号と、その背後にある計測したい概念の真の値の対応関係を reporting function (RF) と呼ぶ。本章では、この RF の形状が不明だが回答者全員同一の場合と、回答者によって異なる場合に分けて、問題を整理する。結論としては、様々な対処方法が検討されているものの、現状、完璧な対処方法はないといえる。順序尺度を用いた分析をする際には、可能な限り様々な対処手法を試して結果の頑健性をチェックするのが、現時点におけるベストな手法であると思われる。

キーワード： 回答スタイル、回答バイアス、順序尺度、reporting function

1. はじめに

質問紙調査を用いて人々の価値観や信念を測り、その回答の集団の平均値を当該集団の「文化」指標として用いる手法は、経済学の論文でよく用いられている(前章参照)。¹ 例えば、オランダの社会心理学者 Geert Hofstede の国民文化の 6 次元は様々な研究で使用されているが、各国の各次元の指標は質問紙調査の回答から作成されたものである。例えば、国単位の個人主義指標は「雇用の安定性があること」「自分のための時間、家族と過ごす時間が十分にあること」「興味のある仕事に従事すること」「家族や友人から尊敬される仕事に就くこと」という 4 質問に対する 5 段階の回答(1:最優先に重視する、2: 非常に重視する、3: 重視する、4: あまり重視しない、5: 殆ど/全く重視しない)から合成されている(Hofstede and Minkov 2013)。²

この例のような回答は順序尺度(ordinal scale)と呼ばれる。この例では、重要度は 1 が最も高く、その後、2, 3, 4, 5 となるにしたがって重要度が低くなる。このように、重要度の順序・大小は分かるが、各選択肢の間に重要度の違いがどれくらいあるのかは明らかでない。³ このため本来、順序尺度の番号をあたかも得点のように見なして、A さんの回答が 1, B さんの回答が 3, C さんの回答が 5 だった場合に、この集団の「文化」は 3 人の回答の番号の平均値である 3.0 $(=(1+3+5)/3)$ であるとすることはできない。しかし、実際には Hofstede の国民文化の指標を含め、多くの場合にこのような処理がなされている。

このように、順序尺度の数値をあたかも得点のように見なして平均値を計算したり、順序ロジットモデルや順序プロビットモデルを当てはめて回帰分析を行ったりした場合に、どのような問題が生じるのか、またそれらの問題にどう対処すればよいのか、本章は既存研究をもとに検討する。⁴ ここで、順序尺度の背後にある真の値(上の例では真の「重要度」)が回答選択肢とともにどのように変化するかを表したものを reporting function (RF) と呼ぶ。⁵ 第 2 節では、回答者全員が同じ形状の RF を持っているものの、その形状が不明な場合に生じる問題と、それへの対処方法を紹介する。第 3 節ではさらに、RF の形状が個人や集団によって異なる場合に生じる問題とそれへの対処方法について検討する。なお、本章は順序尺度を用いた文化の計測を念頭に置いて執筆したものだが、本章の内容は順序尺度一般に当てはまるものである。

¹ 本章における「文化」の定義は前章と同様、「国・地域・世代・民族などで区別される集団に共有された価値観、信念、規範、思考・反応のパターン」とする。

² 具体的な合成方法は Hofstede and Minkov (2013)を参照。各言語の質問調査票は次のウェブサイトからダウンロード可能:<https://geerthofstede.com/research-and-vsm/vsm-2013/> (2024 年 2 月 5 日アクセス)。

³ 順序尺度については例えば、統計データ利活用センターウェブサイト(<https://www.stat.go.jp/dstart/point/seminar/02/3-1.html>, 2024 年 2 月 5 日アクセス)参照。

⁴ 本章では経済学に加え、心理学や政治学の文献も参考になっている。筆者は経済学が専門であるため、経済学以外の文献への理解は不十分である可能性があること、またすべての関連論文を網羅しているわけではないこと、留意いただければ幸いである。

⁵ 今回、訳語を見つけることはできなかったが、訳すとすれば「回答関数」などが妥当だと思われる。

2. Reporting Function (RF)の形状が全員同一だが不明の場合

2.1 問題点

話を分かりやすくするため、「1: 全く重要でない、2: あまり重要でない、3: どちらともいえない、4: ある程度重要である、5: 非常に重要である」という 5 段階の順序尺度を例に話を進めていきたい。この順序尺度の番号と、その背後にある真の重要度の対応関係が RF である。順序尺度の番号(1, 2, 3, 4, 5)に対応する真の重要度が例えば、(1, 2, 3, 4, 5)であれば RF は線形の形状をしており、真の重要度が(1, 3, 4.2, 4.8, 5)のように選択肢の番号が小さいほど選択肢間の重要度の変化が大きければ凹状 (concave) RF、反対に真の重要度が(1, 1.2, 2, 3.2, 5)のように選択肢の番号が大きいほど選択肢間の重要度の変化が大きければ凸状 (convex) RF となる。もちろん、これ以外により複雑な形状の RF もありうる。ここで今、回答者全員が順序尺度の番号を同じように認識している、つまり RF の形状は全員同じであるとするとする。しかし、その RF の形状は不明の場合にどのような問題が生じるだろうか。この点、経済学の分野で近年、研究が盛んである。

経済学では、[Schröder and Yitzhaki \(2017\)](#)や [Bond and Lang \(2019\)](#)が、RF の形状が全員同じでも形状が不明な場合、強い仮定がない限り、順序尺度を用いた集団の平均値の比較は無意味なこと、順序尺度を被説明変数とした最小二乗法 (OLS) 回帰や順序プロビット(またはロジット)回帰の係数の符号は特定できないことを指摘した。当時既に、順序尺度回答から幸福度を測定する研究がかなり蓄積しており、それらの研究の意義を根底から覆すことになりかねない指摘であった。

ここでは、[表 1](#) の例を用いて、順序尺度回答を用いた集団の平均値の比較について検討してみる。まず、A と B という 2 つの集団があり、それぞれ集団を代表する 3 人が X(例えば、家族)の重要度に対する 5 段階尺度の設問に回答した場合を考える。その回答分布は、パネル(a)のとおり、ケース 1 とケース 2 の 2 通りある。ここで、回答選択肢に 1~5 の番号はついているが、その背後にある真の「重要度」は分からない、つまり RF の形状は不明である。このため、パネル(b)のとおり RF の形状について 3 つのパターン(RF が線形、凹状、凸状)を想定してみよう。なお、この RF が線形のケースは、順序尺度の番号をそのまま真の「重要度」を表す数値として使用するもので、一般的によく用いられる手法である。ここで、ケース×RF の形状ごとに集団 A と集団 B の回答者の真の「重要度」の平均値を算出し比較した結果がパネル(c)である。⁶ まず、ケース 1 の結果を見ると集団の平均値は、RF が線形または凸状の場合は A、凹状の場合は B の方が高くなる。つまり、ケース 1 では RF の形状が分からない限り、どちらの集団の方が平均的により X を重要視しているかは判断できないのである。一方、ケース 2 では RF の形状に関係なく集団 B の方が A より平均値が高いという結果になっている。

実は、[図 1](#) が示すように、ケース 2 は集団 A と B の回答者の累積分布が first-order stochastic

⁶ 例えば、ケース 1, RF が凹状のときの集団 B の真の「重要度」の平均値は、 $(3 \times 1 \text{ 人} + 4.8 \times 2 \text{ 人}) / 3 \text{ 人} = 4.20$ と算出できる。

dominance (FOSD, 一次確率優越)の関係にあるケース、ケース 1 は FOSD が満たされないケースである。Schröder and Yitzhaki (2017)や Bond and Lang (2019)は、各集団の累積分布に FOSD の関係が満たされる場合にのみ、順序尺度の番号を用いた平均値の比較が可能であることを示した。図 1 からわかるように FOSD が満たされないケース 1 とは、各集団の累積分布関数が交差している場合である。また表 1 のパネル(a)の回答者の分布に戻ると、ケース 1 では、集団 A では回答が 1 と 5 の両極端の選択肢に、集団 B では 2 と 4 の選択肢に分散している。直感的にも、集団 A と B のどちらの「重要度」が平均的に高いと評価できるか、判断が難しそうなのがわかるだろう。

2. 2 問題への対処・反論

最近になって、Schröder and Yitzhaki (2017)や Bond and Lang (2019)が提起した問題に対処、または反論する論文がいくつか発表されている。まず、順序尺度の集団の平均値を比較するのではなく、肯定的な回答の割合 (Sechel 2021) や中央値 (Chen et al. 2022) で比較することを推奨する論文がある。Chen et al. (2022)は中央値 (median) の比較においては、平均値の比較に必要な上述の FOSD の仮定が不要であることを示した。実際に表 1 のパネル(c)にみられるように、FOSD の仮定が満たされないケース 1 においても中央値での比較であれば、RF の形状に関わらず常に集団 A の方が集団 B よりも X を重要視していると判断できる。また、Bloem and Oswald (2022)は回帰分析の被説明変数として順序尺度を用いる場合、中央値より上なら 1、下ならゼロをとるダミー変数の使用を推奨している。

また、順序尺度の番号の数値をそのまま使用して(つまり、線形 RF を仮定して)、集団間の平均値を比較し回帰分析を行う、従来的一般的手法を擁護する論文も登場した。Kaiser and Venrdik (2023)は、回帰分析の被説明変数として順序尺度を用いる場合、線形 RF であれば回帰係数の符号の判断はできること、かつ過去の実証結果からは多くの場合、線形 RF と判断できることを示した。また、一般的な主観的厚生 (Subjective well-being: SWB) の 11 段階の順序尺度を重要な説明変数に回帰した場合に、係数の符号が逆になることは非常に稀かほとんど不可能であることも示した。Kaiser and Oswald (2022)は、満足度の順序尺度の番号の数値の方が、社会経済的変数群 (絶対所得、相対所得、自営業かどうか、世帯内の大人や子供の数、婚姻状態、子供の誕生有無、雇用状況、企業規模、教育水準、居住地域) よりも、その後の行動 (引っ越し、離婚、離職、病院受診) に対する説明力が高いことを英国、ドイツ、オーストラリアのデータを用いて示し、満足度の順序尺度の番号そのものにも意味があることを主張している。

3. Reporting Function (RF)の形状が個人・集団によって異なる場合

3. 1 問題点

次に、RF の形状が不明で、かつ集団や個人によって異なる場合を考えよう。このとき例えば、表 1 の例で、A さんの RF は線形なのに対し、B さんの RF は凹形であり、A さんも B さんも同じ

回答選択肢 2 を選んだのに、背後にある真の重要度は A さんが 2、B さんが 3 のようなケースを考える。この場合、選んだ選択肢の番号は A さんも B さんも同じなのに、真の重要度は B さんの方が高いため、順序尺度の番号で A さんと B さんの真の重要度を比較することはできないことが分かるだろう。RF の形状が全員同一でも不明な場合に既に問題が生じていたため当然ではあるが、RF の形状が不明かつ集団や個人によって異なる場合にも、順序尺度の番号を用いた集団間の平均値の比較や回帰分析は意味をなさないものになってしまう。

RF の形状が個人や集団によって異なる要因としては、準拠集団 (reference group) の違いや回答バイアス (response bias) などが挙げられている。準拠集団の違いが要因になる状況とは例えば、自分自身の X に対する真の重要度は A さんも B さんも同じなのに、周りに X を重要視する人が多い A さんは周りの人を基準に考えて「4:ある程度重要である」と回答する一方、周りに X を重要視する人が少ない B さんは「5:非常に重要である」と答えるような状況である。このように、準拠集団が異なると、それと自身の状況を比較して回答選択肢を選んでしまう可能性があり、計測したい真の値の比較ができなくなるのである (Heine et al. 2002)。

次に、回答バイアスとは「本来測定したい内容とは異なる、項目への回答にあたっての系統的な傾向」(北條・岡田 2018; Paulhus 1991) である。Wetzel et al. (2016) の分類によると、回答バイアスには、質問項目の内容に関係ない一般的な回答バイアスと、質問項目の内容に依存する回答バイアスの二種類がある。前者には、後述する回答スタイル (response style) や、ハロー効果 (halo effect)、寛大化・厳格化傾向 (leniency/severity bias) といった評定者 (rater) バイアスが含まれ、後者には社会的望ましき反応 (socially desirable responding: SDR) が含まれる。ハロー効果は評価対象の全ての特徴を、観察できていない項目も含め肯定的または否定的に評価する傾向、寛大化・厳格化傾向は全ての評価について寛大または厳格に評価する傾向のことである (Wetzel et al. 2016)。社会的望ましき反応は「回答者の回答を社会的に良いとされる方向に歪ませる反応バイアス」(藤岡 2023) のことである。例えば、社会的に家族を大切にすべきだと考えられている文化では、(特に自分の回答が他者に見られる可能性のある場合に) 自分の真の態度以上に、家族は重要だと回答する可能性が生じる。

回答スタイルについて先行研究では、黙従回答スタイル (acquiescence response style: ARS)、非黙従回答スタイル (disacquiescence response style: DARS)、極端回答スタイル (extreme response style: ERS)、中間回答スタイル (mid-point response style: MRS) が主に挙げられる (北條・岡田 2018)。⁷ 図 2 にみるように、ARS は 5 段階尺度であれば選択肢 4 または 5 を選びやすい、つまり肯定的な回答をしやすい特徴をもち、DARS は反対に選択肢 1 や 2 などの否定的な回答をしやすい特徴をもつ。MRS は選択肢 3 の中間の選択肢を選びやすいスタイルで、ERS

⁷ 他にも、いい加減にランダムに回答する傾向の noncontingent responding (NCR) や、平均付近の狭いまたは広いレンジの選択肢を選ぶ傾向の response range (RR)、相対的に否定より肯定する傾向にある net acquiescence response style (NARS) などがある (Baumgartner and Steenkamp 2001; Van Vaerenbergh and Thomas 2013)。

は反対に選択肢 1 や 5 といった極端な選択肢を選びやすいスタイルである。どの回答スタイルも、設問の内容とは関係なくそのような回答傾向が見られるということが重要である。

文化研究の文脈では、文化圏ごとに異なる回答スタイルが観察されることが指摘されている (Harzing 2006; Johnson et al. 2011; 経済協力開発機構 (OECD) 2015: 187-193; 田崎・申 2017; Bonjeer and Vonkova 2023)。これら研究では例えば、日本、中国、韓国など東アジアの人々が欧米の人々と比べて MRS の傾向があること、スペイン語圏やギリシャは他地域と比べて ERS や ARS の傾向があることなどが見出されている。このように文化研究の文脈で回答スタイルが回答に及ぼす影響は無視できないため、本節では以降、RF の形状が個人や集団によって異なる要因として、主に回答スタイルを念頭において話を進めていく。

集団(または文化圏)ごとに異なる回答スタイルが観察される場合にどのような問題が生じるのか、表 2 の数値例をもとに検討してみよう。ここでは、4 つの集団(文化圏)があり、それぞれ 5 人が各集団を代表して X(例えば、家族)の重要度についての 5 段階の順序尺度の設問に回答している。各人は所属する集団固有の回答スタイル(ERS、MRS、ARS、DARS)を持っている。真の重要度はどの集団も、5 人がそれぞれ 1, 2, 3, 4, 5 の水準にあり、集団平均値はどの集団も 3.0 で等しい(列 4 参照)。本来、これが計測したい結果である。しかし、実際には各集団特有の回答スタイルのせいで、実際の回答(選んだ選択肢)は真の水準からは乖離する。各集団の回答平均値を比較すると、ARS 集団 > MRS 集団 > ERS 集団 > DARS 集団となり、あたかも集団間で「重要度」に違いがあるような結果になってしまう(列 5 参照)。また、回帰分析で「重要度」を所得水準など何かしらの説明変数 Z に回帰した場合にも問題は生じる。表 2 の一番下の行にあるように、本来計測したい真の水準と Z の相関係数は 1.00 であるが、実際の回答と Z の相関係数を求めると 0.41 となり、実際の回答からは本来計測したい真の重要度と Z の関係を正確に推定することはできない。

3.2 問題への対処方法

RF の形状が不明で、かつ集団や個人によって異なる場合の対処方法としては、調査設計時に可能な対処例と、調査後または既存調査利用時の対処例の二種類に分けられる。

3.2.1 調査設計時の対処例

① 調査方法・設問の仕方の工夫

これは最も簡単にできる対処の仕方である。例えば、SDR を避けるために回答者の回答が周囲の人や調査者にわからないよう匿名性を確保する (Paulhus 1991; Johnson et al. 2011); MRS を避けるために「3: どちらでもない」といった中間選択肢をそもそも設けない (Wetzel et al. 2016); MRS に加え ERS も回避できるよう、回答選択肢は 2 つに絞る (Paulhus 1991); ARS や DARS を特定できるよう、ある概念を測るための設問を複数設け、設問の半分はその概念に肯定的な尺度で尋ね、もう半分は否定的な尺度で尋ねるようにする (Paulhus 1991; Johnson et al. 2011; Wetzel

et al. 2016)などがこれに該当する。⁸ また、行動の頻度など、順序尺度以外のより客観的な指標で概念の計測が可能な場合には、そもそも順序尺度を用いないという判断もありうる。⁹

② 回答バイアス測定用設問を設ける

これは、特定の回答バイアスが特に疑われる際に、その回答バイアスの程度を計測するための設問をあらかじめ調査票に盛り込んでおくというものである。例えば、SDR については「社会的望ましき尺度」を計測できる設問群が複数開発されている(藤岡・脇田 2023; Lanz et al. 2023)。ただ、同尺度については、SDR 以外のものを計測している可能性がある、同尺度自体が SDR やその他の回答スタイルに影響されるなどの批判がなされている(Wetzel et al. 2016; Lanz et al. 2023)。

また、SDR とも関連する自己高揚(self-enhance)傾向については、存在しない概念などについての知識を問う overclaiming という手法がある(Paulhus et al. 2003)。OECD が世界各国の 15 歳児を対象に行う学習達成度調査 PISA (Programme for International Student Assessment) の 2012 年調査では、数学の 16 の概念についての知識の程度を 5 段階尺度で尋ねているが、そのうち 3 つは存在しない概念であり、それらの回答から個々人の自己高揚傾向を測定できるようになっている。ただし、overclaiming についても、知識を尋ねる設問に用途が限定される、本当に自己高揚傾向を測っているのか疑問視する意見があるなど、課題がある(He et al. 2020)。

これらの手法は、回答バイアス測定用尺度からきちんと回答バイアスの程度を測れたとしても、質問数が増えるために回答者の負担が増し、回答率が下がったり、いい加減な回答が増えたりする恐れがある。また、尺度で測れない回答バイアスは制御できないという課題がある。

③ パネル調査を行い個人固定効果を制御

同じ個人に対して複数回、同じ設問の調査を行うパネル調査を実施することで、調査結果を回帰分析する際に個人の固定効果をコントロールし、個人固有の RF の形状が結果にもたらす影響を取り除くことが可能となる。この手法の主な問題点は、調査コストがかさむこと、性別や教育水準、居住国など、時間とともに(ほとんど)変化しない個人属性が結果に与える影響は個人固定効果に吸収されてしまって推定できないことである。

④ 係留ビネット法(anchoring vignettes)

係留ビネット法とは: 係留ビネット法は King et al. (2004)によって提唱された手法で、ハーバード大学の政治経済学者 Gary King 教授のウェブサイト¹⁰に、関連する論文やスライド、係留ビネッ

⁸ 他の対処例については、Johnson et al. (2011), Wetzel et al. (2016)など参照。

⁹ 例えば、Falk et al. (2018)が実施した Global Preference Survey で利他性を計測する設問の一つは、ある具体的金額をもらったと想像してもらい、そのうちいくらを寄付するか具体的金額を答えてもらうものである。また、Kahneman and Krueger (2006)は、望ましくない状況に割く時間割合という客観的指標から SWB の計測を試みた。

¹⁰ <https://gking.harvard.edu/vign> (2024 年 2 月 7 日アクセス)

トを使用した調査の例、実践的なアドバイスなどの資料が掲載されている。係留ビネット法とは「回答者が従来の自己評定項目に回答することに加えて、自己評定項目に即した複数人の仮想人物について描写したビネット(短文)を読んだうえで、その仮想人物について評定する項目(ビネット項目)に回答することを求める調査法である」(北條・岡田 2018)。

具体的にどうということなのか、係留ビネット法を使用した [Angelini et al. \(2014\)](#) の論文を例に説明していこう。同論文は、SHARE (Survey of Health, Ageing and Retirement in Europe) の 2006-07 年調査を使用し、欧州 10 カ国の 50 歳以上の市民 5,606 人の人生満足度を分析した論文である。調査参加者は、「あなたは全般的に人生にどれくらい満足していますか? —1: とても不満、2: 不満、3: どちらでもない、4: 満足、5: とても満足」¹¹ という自己評定項目で、自分自身の人生満足度を尋ねられる。一方、調査参加者には 2 つのビネットが提示される。1 つ目のビネットには 63 歳の John という男性、2 つ目のビネットには Carry という 72 歳の女性について、短い描写がある。例えば、John については、2 年前に死別した妻のことをよく考えている、定期的に訪れる 4 人の子供と 10 人の孫に高価なプレゼントをするほどの余裕はない、心臓の問題で最近仕事を辞め、疲れやすいが、深刻な健康問題は抱えていない、といった描写である。この描写を読んだあとに、「あなたは、Jonh が人生にどれくらい満足していると思いますか?」と尋ねられ(ビネット項目)、自己評定項目と同じ 5 段階尺度の回答が求められる。Carry についても同様である。

2 つの仮定: 係留ビネット法では 2 つの重要な仮定が置かれている([King et al. 2004](#))。第一の仮定は「反応一貫性(response consistency)」であり、回答者は自己評定項目とビネット項目への回答に同じ RF を用いて回答しているという仮定である。つまり、自分も John も幸福度の状況は同じくらいだと感じたのに、自分自身の人生満足度については「5:非常に満足」と答え、John については「4: 満足」と答えるといったようなことは起こらないと考える。第二の仮定は「ビネット等価性(vignette equivalence)」であり、どの回答者もビネットを同じように解釈しているというものである。例えば、John から感じる真の幸福度の水準はどの回答者も同じであると考えるのである。この二つの仮定を置くと、回答者間の自己評定項目回答の違いは、「その人の真の人生満足度の違い + 回答バイアスなど RF の違い」で説明できる一方、ビネット項目回答の違いは RF の違いだけで説明できることになり、自己評定項目回答から RF の違いを除去することが可能となる。

パラメトリックな手法: 実際に、どのように RF の違いを除去し、真の人生満足度の個人または国間での比較ができるだろうか。[Angelini et al. \(2014\)](#) を含め多くの論文で使用されているパラメトリックな手法を紹介する前に、[表 3](#) の例を用いて、係留ビネット法の直感的理解につながるノンパラメトリックな手法([King et al. 2004](#))を紹介しよう。[表 3](#) のパネル(a)では、A, B, C 3 人の自己評定項目および Jonh と Carry のビネット項目で選択した回答が掲載されている。3 人とも自己評定項目では「3: どちらでもない」を選んでおり、これを彼らの真の人生満足度だとみなすと、人生満足度は全員同じだということになる。しかし、ビネットに対して選んだ番号は全員異なる。「ビネット

¹¹ 筆者訳。原文は”How satisfied are you with your life in general? —1: very dissatisfied, 2: dissatisfied, 3: neither satisfied, nor dissatisfied, 4: satisfied, 5: very satisfied.”

等価性」を仮定すると、3 人とも John(または Carry)から感じる真の幸福度の水準は同じである。このため、最終的に選んだ番号が異なるのは、同じ幸福度に対して 3 人が異なる番号を対応させている、つまり RF が異なるためと解釈できる。例えば、A さんは ERS、B さんは DARS の傾向、C さんは ARS の傾向があるといった具合である。ここで、John と Carry から感じる真の幸福度は全員同じなのであるから、John と Carry への評価と比較して自分自身がどこに位置しているか、その相対的位置を比べることで、3 人の真の人生満足度を比較することが可能となる。まず、表 3 にあるように、John と Carry への評価との相対的位置を示す新しい 5 段階尺度を作成し、その尺度を用いると、A さんの人生満足度は 3 (Carry と John の間)、B さんは 5 (Carry より上)、C さんは 1 (John より下)となり、真の人生満足度は B さんが最も高く、C さんが最も低いと解釈できるのである。ただ、このノンパラメトリックな手法は、同順位や順序に矛盾があった場合に、それらを包含するような新しい尺度カテゴリーを作る必要が生じる結果、元の細かい情報が無駄になってしまうといった欠点がある (King et al. 2004)。

ノンパラメトリックな手法: 実際に多くの論文で使用されるパラメトリックな手法が CHOPIT (compound hierarchical ordered probit, 直訳すると合成階層順序プロビット)モデルである。通常の順序プロビットモデルでは全員が同じ RF (より具体的には、被説明変数の真の値がどの範囲にあれば順序尺度のどの選択肢を選ぶかという閾値)を持っていると仮定されているが、CHOPIT は人によって閾値が異なることを想定した推定手法である。上述のとおり、ビネット項目への回答から各人の閾値を推定できるため、自己評定項目回答から閾値の違いによる部分を取り除くことができ、その人の真の被説明変数の値や様々な説明変数が与える影響を推定できるのである。¹² 例えば、Angelini et al. (2014)は 50 歳以上人口の人生満足度の欧州 10 カ国の平均値を比較した場合、自己評定項目の回答をそのまま用いた場合にはフランスが下から 3 番目、デンマークが最も高い満足度であったのに対し、CHOPIT により個々人の RF の違いがもたらす影響を除去すると、フランスは上から 3 番目、デンマークは上から 5 番目の満足度になることを見出した。

使用研究・調査例: 係留ビネット法によって個人間の RF の違いがもたらす影響を除去しようとする実証研究は多数ある。人生満足度については Angelini et al. (2014) に加えて、Kapteyn et al. (2010), Angelini et al. (2012), Bertoni (2015), Montgomery (2022)、仕事満足度について Kristensen and Johansson (2008), Bonsang and van Soest (2012)、所得満足度について Bonsang and van Soest (2012), Kapteyn et al. (2013)、主観的健康観について King et al. (2004), Kapteyn et al. (2007), d’Uva et al. (2008, 2011), Molina (2016), Peracchi and Rossetti (2013)、政治的有効性感覚について King et al. (2004)、飲酒について van Soest et al. (2011)、主観的貧困について Ravallion et al. (2016) などがある。また、前述の欧州の健康調査 SHARE の他、OECD の PISA

¹² CHOPIT モデルについては式をみるとより理解が深まる。分かりやすい説明として例えば、Greene et al (2021)や Montgomery (2022)参照。

や世界保健機関(WHO)の World Health Survey (WHS)といった大規模な国際比較調査にもビネット項目が導入されている(北條・岡田 2018)。

長所・短所: 係留ビネット法の長所は、特定の回答スタイルを想定してそれだけに対応するのではなく、個人間の RF の違いをもたらすような全ての要因に対応できる点である。一方、短所はビネット項目を調査票に含める必要があるため、回答者の負担が増すことその他、「反応一貫性」と「ビネット等価性」の仮定が本当に成立しているのか疑問が残る点である。特に「ビネット等価性」はかなり強い仮定だと言えよう。例えば、辛い状況にある人ほど上述の John の状況を幸せな状況だと評価するような、ビネット等価性が満たされない場合も十分に想定できる。

「反応一貫性」と「ビネット等価性」の仮定への批判に対し、これら仮定の成立をテストしようとする論文もある。「反応一貫性」については Gupta et al. (2010)や d'Uva et al. (2011)が、「ビネット等価性」については d'Uva et al. (2011), Angelini et al. (2014), Montgomery (2022)がテストを試みている。また、Peracchi and Rosseti (2013)は両仮定を同時検定(joint test)する手法を提案し、Greene et al. (2021)は CHOPIT を微修正することで両仮定を同時または別々にテストする手法を開発している。

3. 2. 2 調査後の対処例

自身で調査をする際には、前節で述べた調査設計時の様々な工夫を盛り込むことが可能だが、他者が行った調査のデータを使用する際には分析時に対処するしかない。本節では、既に調査を実施してしまった後の結果分析時の対処例を紹介する。

① 二値尺度に変換する

例えば、「1: 全くそう思わない、2: そう思わない、3: どちらでもない、4: そう思う、5: 非常にそう思う」という 5 段階尺度があった場合に、4 と 5 の肯定的回答を 1、それ以外の 1, 2, 3 の回答を 0 に変換した二値変数を分析すれば、ERS の影響を回避することが可能となる。また、完全に MRS の影響をなくすことはできないが、4 と 5 の肯定的回答を 1 とする場合に加えて、3, 4, 5 の回答を 1 に、残りの 1, 2 を 0 に変換するケースも分析して結果に同じような傾向が見られれば、MRS の影響を考慮しても結果は一定程度頑健であると主張することはできるだろう。ただ、二値尺度への変換では、ERS や MRS 以外の回答バイアス(ARS, DARS, SDR など)や、異なる準拠集団によって生じた RF の違いには対応できないという問題がある。

② シンプルな回答スタイル指標を計算しコントロールする

回答スタイルについてレビューした Van Vaerenbergh and Thomas (2013)が、回答スタイルを特定し修正するために推奨する手法の一つが、representative indicators for response styles (RIRS)と呼ばれる手法である。RIRS を使用した論文例としては、Baumgartner and Steenkamp (2001)や Buckley (2009)などがある。この手法は、個人ごとに様々な回答スタイルの指標を計算し、それを

回帰分析の際にコントロールすることで回答スタイルの影響を除去しようとするものである。具体的には、個人ごとに、関心のある変数を構成する質問項目とは異なる多様な質問項目への回答から以下計算する(以下、①で使用した5段階尺度の場合):

ERS 指標=1 または 5 と回答した質問項目の割合

MRS 指標=3 と回答した質問項目の割合

ARS 指標=4 または 5 と回答した質問項目の割合

DARS 指標=1 または 2 と回答した質問項目の割合

ARS、DARS 指標については、より極端な回答(5 や 1)のウェイトを2倍にして割合を計算することもある。¹³ ここで重要なのは、項目間の相関が低い多様な質問項目から回答スタイル指標を作成することで、質問内容に依存しない回答傾向を拾うことができるという点である。

RIRS の長所は理解がしやすく計算が簡単なことである。短所は、多様な質問項目がある調査にしか適用できないこと、また、想定した回答スタイル以外の個人間の RF の違いには対処できないことがある。また、回答スタイル指標を推定してから、その推定値を回帰分析の説明変数として加えると、第一段階の推定における誤差が含まれたまま第二段階を推定することになり、誤差が蓄積するという問題も指摘されている(Weijters et al. 2008)。

③ SEM (Structural Equation Model: 共分散構造分析または構造方程式モデル)

RIRS とともに Van Vaerenbergh and Thomas (2013)が推奨するもう一つの手法が representative indicators response styles means and covariance structure (RIRMACS)と呼ばれるものである。これは RIRS の最後の課題(誤差の蓄積)を克服する手法である。つまり、Structural Equation Model (SEM: 共分散構造分析または構造方程式モデル)によって回答スタイル指標の推定と第二段階の回帰分析の推定を同時に行うことで、より正確な推定が可能となる。使用例としては、Weijters et al. (2008)や田崎・申(2017)がある。ただ、この手法も RIRS と同様、多様な質問項目が必要であること、想定した回答スタイル以外には対応できないという課題がある。

④ IRT (Item Response Theory: 項目反応理論)

項目反応理論(IRT: Item Response Theory)はもともと、テスト理論(test theory)として開発され、難易度や受験者集団が異なるテストを受けた受験者間の潜在的な能力を比較可能にする手法として発達してきた(宇佐美 他 2019)。しかし近年、個人レベルで本来測定したい概念(例:幸福度)と回答スタイルを分離させる手法としての使用が増加している(Falk and Ju 2020)。複数の手法があるようだが、ここでは Adams et al. (2019)等が使用している Multidimensional Item Response Theory (MIRT)モデルを紹介する。

¹³ その他、NCR 指標=高相関の質問項目のペアを多数選び、各ペアの回答の番号の差の絶対値の合計、RR 指標=全回答の番号の標準偏差、として計算されている(Baumgartner and Steenkamp 2001; Buckley 2009)。

今、回答者が R 人、幸福度を測るための質問項目が I 個、全ての質問項目は 5 段階の順序尺度(1: 全くそう思わない, 2: そう思わない, 3: どちらでもない, 4: そう思う, 5: 非常にそう思う)であるとしよう。MIRT では多項ロジットモデルを想定し、回答者 r の質問項目 i への回答 U_{ri} が選択肢 k である確率を以下のように表す:¹⁴

$$P(U_{ri} = k | \theta_r, s_r) = \frac{\exp(a_{ik}\theta_r + s_{rk} + c_{ik})}{\sum_{h=1}^K \exp(a_{ih}\theta_r + s_{rh} + c_{ih})}$$

ここで、 θ_r は回答者 r の潜在特性、つまり本来測定したい真の幸福度を表す。 $s_r = [s_{r1}, s_{r2}, s_{r3}, s_{r4}, s_{r5}]$ は、回答者 r の回答スタイルを表す。 $\sum s_{rk} = 0$ で、 s_{rk} が正に大きいほど選択肢 k を選ぶ傾向があり、負に大きいほど k を選ばない傾向があることを示している。 K は選択肢の数で現在の例では $K=5$ である。例えば、 $s_r = [2, -1.3, -1.3, -1.3, 2]$ の人は、2, 3, 4 の選択肢と比較して、「1: 全くそう思わない」や「5: 非常にそう思う」を選ぶ傾向、つまり ERS の傾向があるといえる。 $a_i = [a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5}]$ は scoring function と呼ばれるもので、質問項目 i の回答選択肢をどう評価するかを表す(または選択肢間の間隔を表しているともいえる)。 a_i は通常、 $a_i = [-2, -1, 0, 1, 2]$ とするなど事前に固定し、かつ全ての項目に同じ a_i が適用される。 $\sum c_{ik} = 0$ で、 c_{ik} は回答者とは関係なく、項目 i について選択肢 k がどれくらい選ばれやすいかを表す($c_i = [c_{i1}, c_{i2}, c_{i3}, c_{i4}, c_{i5}]$)。

上の式で、 $U_{rik}^* = a_{ik}\theta_r + s_{rk} + c_{ik} + \varepsilon_{rik}$ と定義すると(ε_{rik} は誤差項)、 k 以外のすべての選択肢 h について $U_{rik}^* > U_{rih}^*$ が成立すれば、回答者 r は項目 i において選択肢 k を選択する。たとえば今、A, B, C の 3 人がいるとしよう。3 人の違いは回答スタイルだけで、A さんはどの選択肢も均等に選ぶような回答スタイル $s_r = [0, 0, 0, 0, 0]$ 、B さんは ERS で $s_r = [2, -1.3, -1.3, -1.3, 2]$ 、C さんは MRS で $s_r = [-0.5, -0.5, 2, -0.5, -0.5]$ とする。ここでは回答スタイルの違いがもたらす影響を明確にするために、真の幸福度 θ_r は全員 1、 c_{ik} と ε_{rik} はゼロ、 $a_i = [-2, -1, 0, 1, 2]$ としてみる。すると、A さんの $U_{ri}^* = [-2, -1, 0, 1, 2]$ 、B さんの $U_{ri}^* = [0, -2.3, -1.3, -0.3, 4]$ 、C さんの $U_{ri}^* = [-2.5, -1.5, 2, 0.5, 1.5]$ となり、A さんと ERS の B さんは U_{rik}^* の最も大きい選択肢 5 を、MRS の C さんは同様に選択肢 3 を選ぶことになる。このように、 s_r によって個人の様々な回答スタイルを表現できることがわかる。¹⁵

この MIRT モデルにおいて、 s_r 、 θ_r 、 c_i が推定したいパラメーターであるが、これらの大量のパラメーターの推定には、マルコフ連鎖モンテカルロ (Markov chain Monte Carlo: MCMC) 法などが用いられる (Bolt et al. 2014; Adams et al. 2019)。推定の結果、個々人について、回答スタイルと回答スタイルを除去した本来計測したい真の幸福度の推定値を得ることができるのである。事前に回答スタイルを特定する必要がなく、想定もしていないような全ての回答スタイルに柔軟に対処

¹⁴ 推定モデルの説明は Bolt et al. (2014)、北條・岡田(2018)、Adams et al. (2019)を参考にした。特に変数の表記は北條・岡田(2018)に倣った。

¹⁵ 各選択肢を選ぶ確率と潜在特性 θ_r (例: 真の幸福度) の関係を、回答スタイルごとに分かりやすくグラフ化したものとして、Bolt et al. (2014) Figure 2 などを参照されたい。

できることが MIRT モデルの長所である。一方で短所は、個人ごとに s_r や θ_r を推定するため、 θ_r (例: 幸福度)を測るための質問項目が一定数必要となることである。Adams et al. (2019)はシミュレーションの結果、50 超の質問項目が望ましいとする。¹⁶ もともと IRT が応用されてきた学力テストとは異なり、多くの質問項目から幸福度などの一つの概念を測るような調査は回答者の負担も大きく一般的ではないと思われ、そもそも IRT を応用できる調査が少ないのが最大の欠点であると言えよう。

4. 結び

本章では、順序尺度を用いて個人や集団間の平均値を比較したり、回帰分析を行ったりする際に生じる問題およびそれらの問題への対処方法について、既存研究をもとに検討してきた。様々な対処方法が検討されているものの、それぞれ長所・短所があり、完璧な対処方法はないと言える。順序尺度を用いた分析をする際には、可能な限り様々な対処手法を試して、結果の頑健性をチェックするというのが、現時点におけるベストな手法であると思われる。

参考文献

【日本語文献】

- 宇佐美慧, 荘島宏二郎, 光永悠彦, 登藤直弥. 2019.「項目反応理論 (IRT) の考え方と実践」教育心理学年報, 58: 321-329.
- 経済協力開発機構 (OECD) 編著, 桑原進 監訳, 高橋しのぶ 訳. 2015.『主観的幸福を測る: OECD ガイドライン』明石書店.
- 田崎勝也, 申知元. 2017.「日本人の回答バイアス—レスポンス・スタイルの種別間・文化間比較」心理学研究, 88(1): 32–42.
- 藤岡慧. 2023.「社会的望ましさを考慮した項目反応モデルによるパラメタ推定法の提案」心理学研究, 94(4): 300-307.
- 藤岡慧, 脇田貴文. 2023.「公募型 Web 調査データを用いた社会的望ましさ尺度短縮化の試み」関西大学心理学研究, 14 : 53-63.
- 北條大樹, 岡田謙介. 2018.「係留ビネット法による反応スタイルの分類—ヨーロッパの大規模健康調査を例に」行動計量学, 45(1): 13-25.

【英語文献】

- Adams, Daniel J., Daniel M. Bolt, Sien Deng, Stevens S. Smith, and Timothy B. Baker. 2019. “Using Multidimensional Item Response Theory to Evaluate How Response Styles Impact

¹⁶ IRT を適用するには一般的に一定程度の回答者数、質問項目数が必要となる。例えば、DeMars (2010)は 2 または 3 パラメター・ロジスティックモデルにおいて、潜在特性値 θ_r が正規分布に近く識別力が高いなどの条件があれば、回答者 500 人、20 項目あればよいとするが、こういった条件がない場合、回答者 1000 人、40 項目が望ましいとしている。

- Measurement.” *British Journal of Mathematical and Statistical Psychology*, 72(3):466-485.
- Angelini, Viola, Danilo Cavapozzi, Luca Corazzini, and Omar Paccagnella. 2012. “Age, Health and Life Satisfaction Among Older Europeans.” *Social Indicator Research*, 105 (2), 293–308.
- Angelini, Viola, Danilo Cavapozzi, Luca Corazzini, and Omar Paccagnella. 2014. “Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-specific Scale Biases.” *Oxford Bulletin of Economics and Statistics*, 76 (5), 643–666.
- Baumgartner, Hans, and Jan-Benedict E.M. Steenkamp. 2001. “Response Style in Marketing Research: A Cross-National Investigation.” *Journal of Marketing Research*, 38, 143–156.
- Bertoni, Marco. 2015. “Hungry Today, Unhappy Tomorrow? Childhood Hunger and Subjective Wellbeing Later in Life.” *Journal of Health Economics*, 40: 40–53.
- Bloem, Jeffrey R., and Andrew J. Oswald. 2022. “The Analysis of Human Feelings: A Practical Suggestion for a Robustness Test.” *Review of Income and Wealth*, 68(3): 689-710.
- Bolt Daniel M., Yi Lu, and Jee-Seon Kim. 2014. “Measurement and Control of Response Styles Using Anchoring Vignettes: A Model-based Approach.” *Psychological Methods*, 19(4): 528–541.
- Bond, Timothy N., and Kevin Lang. 2019. “The Sad Truth about Happiness Scales.” *Journal of Political Economy*, 127(4): 1629-1640.
- Bonjeer, Tamilka, and Hana Vonkova. 2023. “Relationships Between Response Styles and the Hofstede and GLOBE Dimensions of Culture in a Sample of Adolescents From 33 Countries.” *Cross-Cultural Research*. <https://doi.org/10.1177/10693971231203759>
- Bonsang, Eric, and Arthur van Soest. 2012. “Satisfaction with Job and Income Among Older Individuals Across European Countries.” *Social Indicator Research*, 105(2), 227–254.
- Buckley, Jack. 2009. “Cross-National Response Styles in International Educational Assessments: Evidence from PISA 2006.” New York University, June 2, 2009.
- Chen, Le-Yu, Ekaterina Oparina, Nattavudh Powdthavee, and Sorawoot Srisuma. 2022. “Robust Ranking of Happiness Outcomes: A Median Regression Perspective.” *Journal of Economic Behavior & Organization*, 200: 672–686.
- DeMars, Christine. 2010. “Requirements.” In Christine DeMars. *Item Response Theory*. Oxford University Press. pp.31-37.
- d’Uva, Teresa Bago, Eddy Van Doorslaer, Maarten Lindeboom, and Owen O’Donnell. 2008. “Does Reporting Heterogeneity Bias the Measurement of Health Disparities?” *Health Economics*, 17 (3): 351–375.
- d’Uva, Teresa Bago, Maarten Lindeboom, Owen O’Donnell, and Eddy Van Doorslaer. 2011. “Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity.” *Journal of Human Resources*, 46 (4): 875–906.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018.

- “Global Evidence on Economic Preferences.” *The Quarterly Journal of Economics*, 133(4): 1645–1692.
- Falk, Carl F., and Unhee Ju. 2020. “Estimation of Response Styles Using the Multidimensional Nominal Response Model: A Tutorial and Comparison with Sum Scores.” *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00072>
- Greene, William H., Mark N. Harris, Rachel J. Knott, and Nigel Rice. 2021. “Specification and Testing of Hierarchical Ordered Response Models with Anchoring Vignettes.” *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(1): 31–64.
- Gupta, Nabanita Datta, Nicolai Kristensen, and Dario Pozzol. 2010. “External Validation of the Use of Vignettes in Cross-country Health Studies.” *Economic Modelling*, 27(4): 854-865.
- Harzing, Anne-Wil. 2006. “Response Styles in Cross-national Survey Research: A 26-country Study.” *International Journal of Cross Cultural Management*, 6(2): 243-266.
- He, Jia, Joanne M. Chung, and Fons J. R. van de Vijver. 2020. “An Examination of Different Scale Usage Correction Procedures to Enhance Cross-Cultural Data Comparability.” *Journal of Cross-Cultural Psychology*, 51(10), 794-813.
- Heine, Steven J., Darrin R. Lehman, Kaiping Peng, and Joe Greenholtz. 2002. “What’s Wrong with Cross-cultural Comparisons of Subjective Likert Scales?: The Reference-group Effect.” *Journal of Personality and Social Psychology*, 82(6): 903-918.
- Hofstede, Geert, and Michael Minkov. 2013. “VSM 2013: Values Survey Module 2013 Manual.”
- Johnson, Timothy P., Sharon Shavitt, and Allyson L. Holbrook. 2011. “Survey Response Styles Across Cultures.” In David Matsumoto, and Fons J. R. Van de Vijver eds. *Cross-Cultural Research Methods in Psychology*. Cambridge: Cambridge University Press.
- Kahneman, Daniel, and Alan B. Krueger. 2006. “Developments in the Measurement of Subjective Well-Being.” *Journal of Economic Perspectives*, 20(1): 3-24.
- Kaiser, Caspar, and Andrew J. Oswald. 2022. “The Scientific Value of Numerical Measures of Human Feelings” *PNAS*, 119(42): e2210412119.
- Kaiser, Caspar, and Maarten C.M. Vendrik. 2023. “How Much Can We Learn from Happiness Data?” March 2023.
- Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2007. “Vignettes and Self-Reports of Work Disability in the United States and the Netherlands.” *American Economic Review*, 97 (1): 461–473.
- Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2010. “Life satisfaction.” In Ed Diener, Daniel Kahneman, and John Helliwell eds. *International Differences in Well-Being*. Oxford University Press, New York, pp.70–104.
- Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2013. “Are Americans Really Less Happy with

- Their Incomes?" *Review of Income and Wealth*, 59(1): 44-65.
- King, Gary, Christopher J.L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research." *American Political Science Review*, 98: 191–207.
- Kristensen, Nicolai, and Edvard Johansson. 2008. "New Evidence on Cross-country Differences in Job Satisfaction Using Anchoring Vignettes." *Labour Economics*, 15(1): 96–117.
- Lanz, Lukas, Isabel Thielmann, and Fabiola H. Gerpott. 2023, "Are Social Desirability Scales Desirable? A Meta-analytic Test of the Validity of Social Desirability Scales in the Context of Prosocial Behavior." *Journal of Personality*, 90: 203–221.
- Molina, Teresa. 2016. "Reporting Heterogeneity and Health Disparities Across Gender and Education Levels: Evidence from Four Countries." *Demography*, 53(2): 295–323.
- Montgomery, Mallory. 2022. "Reversing the Gender Gap in Happiness." *Journal of Economic Behavior and Organization*, 196: 65-78.
- Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman eds., *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press, pp.17-59.
- Paulhus, Delroy L., Peter D. Harms, M Nadine Bruce, and Daria C Lysy. 2003. "The Over-claiming Technique: Measuring Self-enhancement Independent of Ability." *Journal of Personality and Social Psychology*, 84(4), 890–904.
- Peracchi, Franco, and Claudio Rossetti. 2013. "The Heterogeneous Thresholds Ordered Response Model: Identification and Inference." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 176(3), 703–722.
- Ravallion, Martin, Kristen Himelein, and Kathleen Beegle. 2016. "Can Subjective Questions on Economic Welfare Be Trusted?" *Economic Development and Cultural Change*, 64(4): 697-726.
- Schröder, Carsten, and Shlomo Yitzhaki. 2017. "Revising the Evidence for Cardinal Treatment of Ordinal Variables." *European Economic Review*, 92: 337-358.
- Sechel, Cristina. 2021. "The Share of Satisfied Individuals: A Headcount Measure of Aggregate Subjective Well-being." *Journal of Economic Behavior & Organization*, 186: 373–394.
- Van Soest, Arthur, Liam Delaney, Colm Harmon, Arie Kapteyn, and James P. Smith. 2011. "Validating the Use of Anchoring Vignettes for the Correction of Response Scale Differences in Subjective Questions." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 174(3): 575–595.
- Van Vaerenberg, Yves, and Troy D. Thomas. 2013. "Response Style in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies." *International Journal of Public*

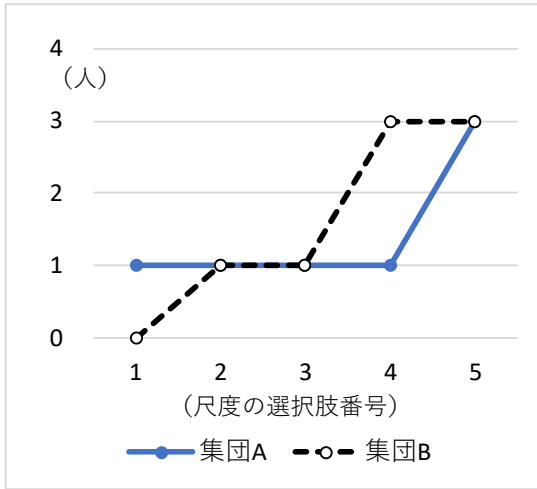
Research, 25(2), 195–217.

Weijters, Bert, Niels Schillewaert, and Maggie Geuens. 2008. “Assessing Response Styles Across Modes of Data Collection.” *Journal of Academy of Marketing Science*, 36, 409–422.

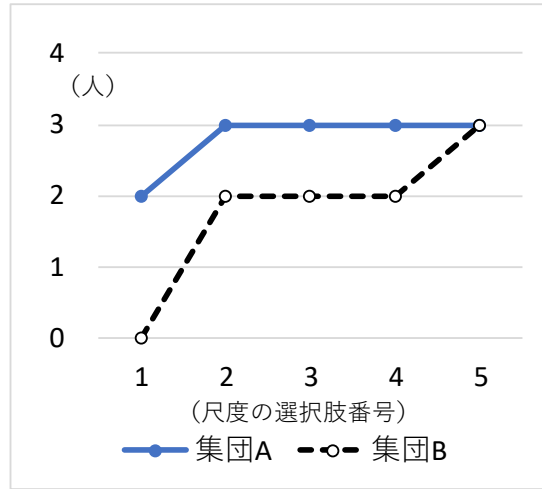
Wetzel, Eunike, Jan R. Böhnke, and Anna Brown. 2016. “Response Biases.” In Frederick T. L. Leong, Dave Bartram, Fanny Cheung, Kurt F. Geisinger, and Dragos Iliescu eds. *The ITC International Handbook of Testing and Assessment*, Oxford University Press, pp. 349-363.

図1. 表1の集団A・Bの回答者の累積分布

(a) ケース1: FOSD 満たさないケース



(b) ケース2: FOSD 満たすケース



注:FOSDはfirst-order stochastic dominance(一次確率優越)の略。

出所:筆者作成

図2. 主な回答スタイル:5段階尺度の例

	1:全く重要でない	2:あまり重要でない	3: どちらともいえない	4:ある程度重要である	5: 非常に重要である
ARS (黙従RS)	○	○	○	●	●
DARS (非黙従RS)	●	●	○	○	○
ERS (極端RS)	●	○	○	○	●
MRS (中間RS)	○	○	●	○	○

注:RSはresponse style(回答スタイル)。黒丸(●)は当該回答選択肢を選びやすい傾向があることを示す。

出所:北條・岡田(2018)、Van Vaerenbergh and Thomas (2013)を参考に筆者作成。

表 1. 順序尺度を用いた集団 A・B の真の「重要度」の比較例

(a) 集団 A・B における回答者の分布

		1:全く重要でない	2:あまり重要でない	3: どちらともいえない	4:ある程度重要である	5: 非常に重要である
ケース 1	集団 A	1 人				2 人
	集団 B		1 人		2 人	
ケース 2	集団 A	2 人	1 人			
	集団 B		2 人			1 人

(b) 順序尺度の背後にある真の「重要度」の数値例

	1:全く重要でない	2:あまり重要でない	3: どちらともいえない	4:ある程度重要である	5: 非常に重要である
RF: 線形	1	2	3	4	5
RF: 凹状 (concave)	1	3	4.2	4.8	5
RF: 凸状(convex)	1	1.2	2	3.2	5

(c) ケース 1, 2 の集団 A・B の真の「重要度」の平均値、中央値の比較

		平均値			中央値		
		A	B	AB どちらが重要視?	A	B	AB どちらが重要視?
ケース 1	RF: 線形	3.67	3.33	A	5.00	4.00	A
	RF: 凹状	3.67	4.20	B	5.00	4.80	A
	RF: 凸状	3.67	2.53	A	5.00	3.20	A
		平均値			中央値		
		A	B	AB どちらが重要視?	A	B	AB どちらが重要視?
ケース 2	RF: 線形	1.33	3.00	B	1.00	2.00	B
	RF: 凹状	1.67	3.67	B	1.00	3.00	B
	RF: 凸状	1.07	2.47	B	1.00	1.20	B

注: X(例: 家族) に対する重要度を 5 段階尺度で尋ねた質問を想定。各集団にはそれぞれを代表する 3 人の回答者がいるケース。

出所: 筆者作成

表 2. 回答スタイルの問題把握のための数値例

(1) 個人 ID	(2) 集団の 回答スタイル	(3) Z	(4) y: 真の重要度	(5) Y: 重要度について の実際の回答 (5段階尺度)
1		10	1	1
2		20	2	1
3	ERS (極端 RS)	30	3	1
4		40	4	5
5		50	5	5
6		10	1	3
7		20	2	3
8	MRS (中間 RS)	30	3	3
9		40	4	3
10		50	5	3
11		10	1	4
12		20	2	4
13	ARS (黙従 RS)	30	3	5
14		40	4	5
15		50	5	5
16		10	1	1
17		20	2	1
18	DARS (非黙従 RS)	30	3	1
19		40	4	2
20		50	5	2
	ERS 集団平均		3.0	2.6
	MRS 集団平均		3.0	3.0
	ARS 集団平均		3.0	4.6
	DARS 集団平均		3.0	1.4
	Z との相関係数		1.00	0.41

出所:筆者作成

表3 係留ビネット法:ノンパラメトリック手法の例

(a) 人生満足度

	Aさん回答	Bさん回答	Cさん回答
5: とても満足	Carry		Carry
4: 満足			John
3: どちらでもない	Aさん自身	Bさん自身	Cさん自身
2: 不満		Carry	
1: とても不満	John	John	

(b) 自身の人生満足度

元の回答	3	3	3
以下の新尺度使用 (ビネットでRFの違い除去)	3	5	1

新尺度:

- 5: 自身が Carry より上に位置
- 4: 自身が Carry と同じ位置
- 3: 自身が Carry と John の間に位置
- 2: 自身が John と同じ位置
- 1: 自身が John より下に位置