

評定尺度法における等間隔性の問題について

— 項目反応理論を用いた検討 —

脇田 貴文

【問題と目的】

心理学の研究において評定尺度法は頻繁に用いられる手法である。評定尺度法では、各カテゴリに一定の得点をあたえ、下位尺度ごとにその得点を合計した尺度得点を用いてさまざまな検討をすることが一般的である。例えば、5件法の場合には、各カテゴリに1点、2点、…、5点という得点を与える。しかし、その前提として各選択カテゴリが等間隔であるかどうかの問題となる。つまり、得られたデータが間隔尺度水準であるか、順序尺度水準であるかということである。このことは後の分析方法・結果にまで影響を与えるという点においても重要である。具体的には間隔尺度水準であれば、総和をとって尺度得点とすること、平均を算出することが可能であり、使用可能な分析方法も多い。しかし、等間隔性が成り立っておらず、順序尺度水準以下であるとすれば、それらの分析方法の使用に疑問が生じる。この問題は、評定尺度の等間隔性の問題に帰着される。現在この問題に対しては、間隔尺度水準とみなしても経験的には大過はないとされている。ただし、「あてはまる」「ややあてはまる」などの評定尺度表現間の間隔を等間隔にする必要があるともいわれ、尺度表現を決める際には織田(1970)の研究が参考とされている。

1970年代までは、評定尺度法に関する研究も内外でいくつか見られたが、その後、このような問題を扱った研究は減少している。その理由としては、一定の成果が得られたためでもあるが、そこには方法的な限界があったとも考えられる。そこで、本研究では、新たに評定尺度の間隔を評価する方法を提案し、どのような場合に等間隔性が成立しているか、また、成立していない場合にどのような影響があるかに関して検討することを目的とする。

【分析モデルと間隔の評価理論】

本研究では、評定尺度における等間隔性の問題を検討するため新たに項目反応理論(Item Response Theory; IRT)のGeneralized Partial Credit Model(Muraki, 1992)を利用した方法を提案し、さまざまな尺度の分析を行う。分析にはPARSCALE(Muraki & Bock, 1993)を用いた。

IRTを利用し、そこにいくつかの仮定をおくことで

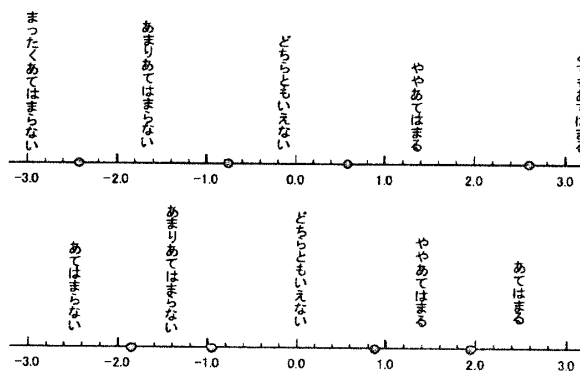
各カテゴリの幅を求めることが可能となる。評定尺度の間隔を評価する場合には、そのカテゴリの幅をその評価の対象とし、各カテゴリの幅が等しければ、評定尺度の等間隔性が成立しているとみなす。

【研究1】

目的 同一の尺度項目で評定尺度表現のみが異なる場合、先に提案した方法により、評価した評定尺度の間隔が変化するか、またどのような影響があるかを2つのデータセットを用いて比較検討することを目的とする。

方法 使用尺度 Rosenberg(1965)のself-esteem尺度10項目の邦訳版。評定尺度表現 データセット1:「まったくあてはまらない」「あまりあてはまらない」「どちらでもない」「ややあてはまる」「とてもあてはまる」の5件法。データセット2:「あてはまらない」「あまりあてはまらない」「どちらでもない」「ややあてはまる」「あてはまる」の5件法。調査対象 データセット1:大学生・短期大学生、女性418名。データセット2:大学生、女性377名。

結果と考察 新たに提案した評価理論を用いて分析を行った。その結果の例として、非逆転項目についての結果をFigure 1に示す。2つのデータセットの違いは、尺度表現に「まったく」、「とても」という修飾語が用いられているかどうかであった。これらの強意の修飾語が用いられることで、評定尺度の間隔が広がったと考えられる。この比較から、尺度表現が、各カテゴリの幅、つまりは評定尺度の等間隔性に影響することが確認された。この



データセット1(上) & データセット2(下)の間隔

Figure 1 データセット1および2における間隔

例のみで結論づけることはできないが、本研究において提案した方法により妥当な結果が得られた。

【研究2】

目的 評定尺度の等間隔性に影響する要因として、項目内容、評定尺度表現、尺度形式、回答者が成人であるか子どもであるか、などが想定される。研究2では、40あまりの尺度をこれらの観点から分類し、その評定尺度の間隔について検討することを目的とする。

結果と考察 全体を通して、評定尺度の等間隔性に対しては、尺度表現、尺度内容の影響が大きいという結果が得られた。特に、尺度項目の内容がポジティブかネガティブかにより評定尺度の間隔の傾向が異なるという興味深い結果が得られた。

【研究3】

目的 研究2では、さまざまな要因により厳密な比較を行うことができなかったため、同一の尺度を用いて、評定尺度表現および尺度形式を変えた9つのバージョンの評定尺度パターンを構成し比較を行う。4件法、5件法、7件法による影響、尺度表現の違いによる影響、尺度形式の違いによる影響、中性カテゴリーの影響について検討することを目的とする。

方法 使用尺度 YG 性格検査の一般的活動性の10項目、神経質の10項目、および、和田(1996)による Big Five 尺度より、情緒不安定性の12項目、外向性の12項目の計44項目を用いた。手続き あらかじめ、9つのパターンの質問紙をランダムにしておき、調査協力者に配布した。調査協力者 大学生2071名(男性660名、女性1371名、不明40名)。

結果と考察 主な結果として7件法は評定尺度の等間隔性が成り立ちにくいこと、また、尺度表現の違い、尺度形式の違いによる回答への影響が少なからず存在することが確認された。また、尺度内容がポジティブかネガティブかにより、「どちらでもない」という中性カテゴリーのもつ意味が異なる可能性が示唆された。

【研究4】

目的 研究1から研究3にかけて、評定尺度の等間隔性についてさまざまな観点から考察した。その結果、実際に使用される評定尺度において等間隔性が成り立っていることが少なく、評定尺度の間隔にはさまざまな要因が影響することが示された。そこで、研究4として、評定

尺度の間隔が分析結果にどのような影響を与えるかについて検討を行う。従来用いられている下位尺度の尺度得点とは異なり、IRTでは、得られた反応が間隔尺度であるということは前提とせず、あくまで順序尺度として扱うことが可能である。したがって、等間隔性の問題を回避することができる。そこで、IRTにおける潜在特性値(以下、 θ とする)と尺度得点を比較する。

方法 (1)中谷(1996)による社会的責任目標尺度の2つの下位尺度を用い、等間隔性がある程度満たされている場合とそうでない場合に、分析結果にどのような影響があるかを比較した。(2)等間隔性という意味では不十分な7件法の例として、達成動機尺度の2つの下位尺度の検討を行った。

結果と考察 (1) θ と尺度得点との比較した場合、等間隔性が成立していない尺度では成立している尺度に比べて、ほぼ同一の θ の回答者であっても、尺度得点では大きな差が生じていた。ただし、基本統計量およびt値に関してはほとんど影響がなかった。(2)7件法に関しては、等間隔性は成り立っていないものの、 θ と尺度得点の間のズレは比較的小さなものであった。

【総合考察】

本研究で新たに提案した方法を用いて、さまざまな尺度を分析し、評定尺度の間隔には以下に示す要因が影響することが明らかになった。尺度表現に関しては、評定尺度の間隔に与える影響は大きく、どのような尺度表現を用いるかは尺度構成上、重要であるという示唆が得られた。また、尺度内容に関しては、項目内容がポジティブな場合と、ネガティブな場合で評定尺度の間隔が大きく異なることが明らかとなった。この知見は、尺度に逆転項目を含めるべきかどうかという問題を考慮する際に参考となると考えられる。本研究で得られた結果は、さまざまな制限がついていたが、評定尺度法に関していくつかの有益な示唆が得られた。

【今後の課題】

評定尺度法は、実施が簡便な反面、その尺度の実施法・構成等により結果が影響を受けやすく、尺度開発および実施の際には、本研究で指摘した点に留意することが必要である。今後、より多くの尺度を分析し、さまざまな視点から検討することで、評定尺度法という手法そのものに関する知見を蓄積していくことが重要である。