リッカート型項目データの間隔データとしての使用

村尾　博

1．はじめに

社会科学の分野におけるアンケート調査では次のようなタイプのデータを作ることが多い。幾らかの意見項目を提示し、各意見項目に対する賛否の程度を被験者（回答者）に聞く。回答選択肢として「強い反対」「中程度の反対」「弱い反対」「どちらでもない」「弱い賛成」「中程度の賛成」「強い賛成」といったように数個の反応カテゴリーを準備しておき、いずれか一つを選択するように調査デザインを設定しておく。そして賛否の順位に従って反応カテゴリーに点数をつける。例えば「強い反対」に－3をつけ、順次－2, －1, 0, ＋1, ＋2と点数化してゆき、「強い賛成」に＋3をつける。「強い反対」に1をつけ、順次2, 3, 4, 5, 6とし、「強い賛成」に7をつける場合もある。反応カテゴリーの数が７個や５個となっている調査デザインが一般的であるが、３個や４個のものも存在する。７個の反応カテゴリーの場合であれば上に示した２つの配点が一般的であり、５個の反応カテゴリーの場合であれば－2から始まり＋2で終わる配点や、1から始まり5で終わる配点が一般的である。このようにして作られるデータが本稿における議論の対象である。

このような点数化は評定尺度法とよばれるが、1932年にレンシス・リッカート（Rensis Likert）が提案したリッカート尺度との類似から、これらを含めリッカート尺度とよばれている。リッカート（1932）が提案した尺度は、測定しようとする量的特性に関連した20個程度の意見項目を用い、被験者ごとに意見項目の点数を合計ないしは平均化し、それを当該量的特性に関する被験者のスコアとするものである。実際には合計点を用いることが一般的であるから、リッカート尺度以外の呼び名としては「相加評定尺度」や「加算評定尺度」が知られている。その一方で、単独の意見項目で測定するものもリッカート尺度と呼ばれている。したがって前者のデータも、後者のデータも、共にリッカート・データと呼ばれている。しかし、両者のデータは全く異なっている。前者のデータは様々に異なった数値からなるが、後者のデータの方は反応カテゴリーの数だけの変動しか見られない。これらを区別するため、本稿の議論の対象である後者のデータは「リッカート型項目データ」と呼ぶことにする。いずれにせよ、冒頭で述べたようなリッカート式のアンケート調査は、数多くの研究で用いられ、現在も用いられている。その領域はリッカートが対象とした態度測定のみならず、パーソナリティ・能力・価値観などの測定にも及んでいる。

リッカート型項目データをデータ分析に用いる研究は数多く存在する。ある学術ジャーナルに発表された188件の研究のうち、その半分以上の95件の研究においてリッカート型項目データが分析に使われている（Clason and Dormody 1994)。リッカート型項目データを分析に使う研究の中には、何のためらいもなく量的データとして分析に用い、平均1.234、標準偏差0.456、相関係数0.678といった数値が報告されている。何のためらいもなく量的データとして用いることは、平均や分散の計算のみならず、相関係数や回帰などの統計的推論にも見られる。

多くの統計学の教科書は、このようなデータは順位データであり、「どちらでもない」と「弱い賛成」との間の違いが、「弱い賛成」と「中程

度の賛成」との間の違いと同等であると仮定できないことを喚起している。評定尺度には順序関係が存在するが、点数の間隔がどこでも同等であるという保証がない。点数の間隔がどこでも同等であるという保証あれば、等間隔の点数化から得たデータを間隔データとして取り扱うことが可能である。しかし、そのような保証がない場合は順位データとして取り扱うべきである。順位データであるならば、足したり割ったりする演算をしても、何ら意味のある情報にはならない。さらに多くの統計学の教科書は、リッカート型項目データが順位データであることを説明した上で、順位データの分析手法も併せて説明している。記述統計に焦点を当ててみると、データの代表値は平均ではなくモードや中央値を用い、データの広がりは標準偏差ではなく四分位偏差や範囲を用いることが紹介されている。

このようなことが統計学の教科書に記載されているにも関わらず、間隔データとして取り扱うことの妥当性を示すことなく、リッカート式項目データを多くの研究者が間隔データとして用い、足したり割ったりするのは何故であろう。他人が量的データとして使っているから自分も使ってよいと思っているケースが一般的であろう。自分の取り扱っているデータが順位データなのか間隔データなのかを検討しても、実用的な判断基準がないので迷うだけであり、迷った末に他人がやっていると同様な方法で間隔データとして用いているケースも考えられる。このような議論は様々あるから自分にとって都合のよい選択をすればよいと考えているケースもあろう。さらには順位の違いがどこでも同等であると仮定して順位データを間隔データとして用いているケースもあるが、単に仮定するだけであり、その仮定の根拠が示されていないのが普通である。順位データを間隔データの如くに取り扱うのであれば、それなりの根拠が示されるべきである。

それでは具体的な事例を紹介しておこう。最初の事例は、統計学に関するさまざまな質問に対し、大学で統計学を教えている青木（2000）が答えるといった内容の電子メールのやり取りであり、それらを「統計学関連なんでもあり」といった形でまとめ、ウェブサイトで掲示している。その中に「多変量分散分析をカテゴリカルなデータで行えるか？」というセクションがあり、ある質問者（学生）は量的なデータなら多変量分散分析で処理できるが、自分が得たデータは５段階の順位データなので多変量分散分析ができないと思っている。しかし、ある先生が「いいんじゃないの」と言ったので、リッカート尺度のようなカテゴリカルなデータで多変量分散分析を行えるか自分でも判断しかねている。それを青木に質問している。青木（2000）は次の３つのアプローチを示している。順位尺度を便宜的に間隔尺度とみなして通常の多変量解析手法を使うアプローチ、順位尺度をそのまま使って数量化理論を適用するアプローチ，順位尺度をダミー変数に変換して通常の多変量解析手法を使うアプローチ。しかし、順位尺度を間隔尺度とみなすアプローチでは、どのような条件のときにそれが可能であるかは示されていない。

第２の事例もウェブサイトで見つけたものであり、リッカート尺度の点数化に関する質問に対し、ハワイ大学のBrown（2000）が回答するといった内容である。その質問は４個の反応カテゴリーに（1, 4, 6, 10）の配点をしても良いかという内容である。これに対する回答は、調査の内容がよく分からないから回答が難しいとしながらも、（1, 4, 6, 10）の配点では連続型の尺度と見なせないことや順位を示す数値になっていないことを指摘し、典型的な（1, 2, 3, 4）の配点が良いとしている。このような配点であれば、名目尺度・順位尺度・連続尺度のいずれかとして分析に使用可能であるとしている。しかし、典型的な（1, 2, 3, 4）の配点を連続尺度のデータとして使用したい場合、どのような条件のときに可能であるかは示されていない。

これらの事例から次のようことが読み取れる。まず、リッカート式アンケート調査に用いる点

数化の基準が曖昧である。さらに数個の反応カテゴリーから得る順位データが、どのようなときに間隔データの如くに取り扱うことができるのかの基準も曖昧である。

付随する問題として次のようなことがある。間隔データとして取り扱うことが正当化できるような場合でも、リッカート式項目データは本質的に誤差を含んでいるという認識は少ないようだ。元のデータが数通りの整数値からなるのにもかかわらず、平均や標準偏差が小数点２桁や３桁まで報告されていることからも、そのことがうかがえる。リッカート式項目データは次のような誤差が内包されたものになることが容易に想像できる。例えば「どちらでもない」と「弱い賛成」との間の違いが、「中程度の賛成」と「強い賛成」との間の違いと同じでないにもかかわらず、等間隔で点数化することから生じる誤差が考えられる。つまり、尺度の歪みから生じる誤差が考えられる。このような誤差を含む情報に基づいて推論するわけであるから、ラフな推論しかできない。このようなことが理解されていない。

リッカート尺度に関連した研究としては、リッカート型調査を実行する際の技術的な問題に取り組むもの、リッカート型調査のデータを分析する際の技術的な問題に取り組むものなど、様々なものがある。その中には測定された内容が意図したものとどれだけ一致しているかと言った妥当性に関し、リッカート（1932）の尺度法を他の尺度法と比較する研究（例えばRoberts 1999）もある。このような学術論文のみならず様々な教科書において、リッカート尺度が説明されている。その多くはリッカート尺度構成法の具体的な手順を説明しているが、どのようにして順位の情報から量的な情報を得ることができるのかと言った肝心な点が説明されていない。したがってリッカート式の簡便な点数化によって得たデータがどのような条件のときに間隔データとして使えるのか不明である。

一方、順位データを間隔データの如くに取り扱うことに関する研究も、いろいろなタイプがある。順位データを間隔データとして取り扱うための条件として正規分布を要求するもの（Borgatta and Bohrnsted 1980）がある一方、ほぼ無条件で順位データを間隔データの如くに取り扱うことができるとするもの（Labovitz 1967, 1968, 1970, 1971, 1972, 1975）もある。良く引用され、かつ他の研究者に大きく影響を与えているのはLabovitz（1967, 1970）である。順位と点数とが単調関数の関係にあることを条件とし、点数化の違いがピアソンの相関係数や平均の差の検定に影響をほとんど与えないことを示し、順位に従って任意に点数をつけた順位データを間隔データの如くに取り扱うことは正当であり、かつ有用であるとしている。このようなLabovitzの研究を受け、Kim（1975）は順位データを順位データとして取り扱うノンパラメトリック手法と、順位データを間隔データとして取り扱うパラメトリック手法とを多変量解析で比較し、パラメトリック手法の方に優位性があるとしている。またGolden and Brockett （1987）は、システマティクな点数化の方法を幾らかのタイプに分類し、それらの違いが与える影響を様々な側面で観測している。それには回答の信頼性への影響、最小２乗回帰への影響、ロジット回帰への影響などが含まれる。

Labovitzの結論に対しては、理論的な側面と実証的な側面との両面から批判が展開されている。批判的な陣営にはChampion（1968)、Morris（1968)、Mayer（1971)、Scweitzer and Scweitzer（1971)、Vargo（1971)、Wilson（1971)、Henkel（1975）などが含まれる。そこでは区間推定の場合、２極化された点数化の場合、多次元の確率変数の場合などでは、Labovitzのような結論が得られないケースがあることが指摘されている。さらに点数化の違いの影響が相関係数に現れにくいことを示しているだけであり、そのことから順位データを間隔データの如くに取り扱うような正当性が得られないこと、統計的な仮定を大幅に崩さないことが大切であることなど、様々なこ

とが指摘されている。当然のことながらLabovitzは自分の結論を擁護する議論を展開している。

このような両陣営の論争から離れ、Labovitz式の点数化をリッカート式の点数化と対比させると、Labovitz式の点数化の実像や欠陥が良く見えてくるようだ。その第１の点は順位データを間隔データの如くに取り扱うことが正当化できるような条件が満たされているか否かの検証に関するものである。Labovitz式の点数化では被験者の態度を真に特徴づける配点と実際に用いる配点の関係が不明のままであり、実際に用いる配点が妥当なものであるかの保証がない。第２の点は個々の被験者から見た場合の議論がなされていない。Labovitz式の点数化では同じ被験者の評価が10点満点中２点でも９点でも良いといったことになる。このようなものが、ものを測る尺度と呼べるか疑問である。このような基本的な観点から、Labovitz式の点数化の欠陥を分かりやしい形で指摘できそうである。一番の大きな問題は、Labovitzを引用することで、直ちに順位データを間隔データとして使って分析を行う研究者がいることである。さらにはLabovitzを引用することもなく、何のためらいもなく順位データを間隔データとして使って分析を行う研究者もいる。したがってLabovitz式の点数化の限界や欠陥を分かりやすい形で示すことは重要であると考える。

本稿の第１の目的は、順位の情報から量的な情報を得るメカニズムがどのようなものであるかをリッカート（1932）の論理から再構築し、その観点からリッカート式項目データを間隔データの如くに取り扱うことができる場合と、そうでない場合との区別を論理的に明確にする。具体的にはモデル化を通じてリッカートの論理を明確にする。リッカートの論理がモデルの形で示されていないことが、さまざまな解釈・誤解を呼び、上で述べたような現象になっていると思われるからである。第２の目的は、リッカート式項目データを間隔データとして用いる場合に、どのような誤差が含まれているかを考察する。このような誤差の分析はリッカート式の簡便な点数化がどのようなものあるかを理解するのに役立つ。さらに、リッカート式の点数化とLabovitz式の点数化とを比較するときにも役立つ。第３の目的は、順位に従って任意に点数をつけた順位データをほぼ無条件で間隔データの如くに取り扱うことを正当化する議論に対し、批判的な分析を行う。具体的にはリッカート式の点数化とLabovitz式の点数化とを比較し、類似点や相違点を調べる。そこからLabovitz式の点数化の欠陥が分かりやすい形で示される。それは容易に理解できる簡単なことであり、かつ重要な意味をもつ。

これ以降のセクションは次のような構成になっている。セクション２ではリッカート（1932）の論理に焦点を置きつつリッカート尺度について説明する。既に述べたことであるが、幾らかの意見項目を合成して作るデータも、単独の意見項目から作るデータも、共にリッカート・データと呼ばれているが、これらは全く異なる。このようなことをセクション３で詳しく説明する。そしてリッカート式項目データを間隔データの如くに取り扱うことが正当化できる場合の条件をセクション４で示す。セクション５ではリッカート式の簡便な点数化に関する工夫について述べる。さらに研究者の間で議論されている反応カテゴリーの数に関する基本的な疑問に対し、リッカートの論理の観点から回答を試みる。セクション６ではリッカート式項目データを間隔データとして用いる場合の測定誤差に関する考察を展開する。この考察を受け、Labovitz式の点数化に関する考察をセクション７で行う。セクション８ではリッカート式の点数化とLabovitz式の点数化とを比較し、Labovitz式の点数化の欠陥を分かりやすい形で示す。本稿のまとめをセクション９で行う。

２．リットカート尺度の説明

まず、リッカート（1932）の論理に焦点を置きつつリッカート尺度について説明する。特に注目する点は、どのようにして順位を示す情報

から量的な情報を作り出すかといった基本的なことにある。

測定しようとしている量的特性は、国際主義に対する態度、黒人に対する態度、帝国主義に対する態度など、直接的に観測できるようなものではなく、いわゆる潜在変数とよばれるものである。このような潜在変数を「測る」ことを目的とし、ひとつの潜在変数に対して多数の意見項目（言説）を準備する。これらの意見項目は潜在変数に関わる具体的な意見を被験者に聞くような内容になっている。ポジティブな態度の意見項目とネガティブな態度の意見項目との双方を準備する。それぞれの意見項目には賛否の程度を数段階に分ける回答選択肢を設けておく。回答選択肢の個数は５個や７個が一般的であるが、調査の内容によって決める。

ある量的特性に関する被験者のスコアを得るまでの過程は大別して次のようになる。

(1) 測定しようとする量的特性に関係する意見項目を多数準備し、アンケート調査を行う。調査結果に基づき、意見項目ごとに被験者のスコアを得る。被験者ごとに全ての意見項目のスコアを加算し、その合計点または平均点を当該量的特性に関する被験者の仮のスコアとする。全ての被験者に対して意見項目の数が同じであることから、合計点を用いるのが普通である。

(2) 幾らかの基準に従って最終的な測定に使う意見項目を選択する。例えばスコアの高い被験者グループと低い被験者グループとの間で、意見が分かれておらず回答にバラツキのないような意見項目は取り除く。さらに測定しようとする量的特性との相関が小さい意見項目も捨てる。具体的には一つの意見項目と他のすべての意見項目との間で相関を計算し、他の意見項目との相関が小さい意見項目は捨てる。このような取捨選択を経て適度に小さいセット（20個か、それよりも少ない個数）になった意見項目を最終的な測定に用いる。

(3) 採用した意見項目のみを用いて被験者の合計点または平均点を計算し、それを当該量的統計に関する被験者のスコアとする。

リッカートはそれぞれ５個の回答選択肢を持つ意見項目を10個程度採用し、最終的な被験者のスコアを得ている。回答選択肢の形式は次のような２つのタイプを用いている。一つのタイプは、意見項目に対して５つの回答文があり、これらの回答文は賛否の程度に従って順位がつく内容になっており、順位に従って配列されているタイプである。彼は「マルチプル・チョイス」型とよんでいる。もう一つのタイプは、意見項目に対して「強い賛成」「賛成」「どちらでもない」「反対」「強い反対」の５つの反応カテゴリーが準備されているタイプである。彼は「強い賛成」型とよんでいる。

このような回答選択肢に点数をつけることになるが、どのような点数化が良いのであろうか。リッカート（1932）が開発した尺度構成法において、重要な点のひとつは正規分布に従う母集団を仮定していることである。左に歪んでいる分布をもつ被験者グループと、右に歪んでいる分布をもつ被験者グループとがある場合は、それらを組み合わせると、分布の歪みを小さくすることが出来ることなどを説明した後、彼は次のように述べている。

> On the basis of this experimental evidence and upon the results of others (8, pp. 542-548, 28, pp. 71-91), it seems justifiable for experimental purposes to assume that attitudes are distributed fairly normally and to use this assumption as the basis for combining the different statements.（リッカート 1932:22）

被験者の散らばりが正規分布に従うと仮定すると、反応カテゴリーにどのような点数をつけたら良いのであろうか。われわれは正規分布に関して次のような知識をもっている。平均μと分散σ^2をもつ正規分布$N(\mu, \sigma^2)$から抽出したデータは、平均μを中心として散らばり、大部分が-3σから$+3\sigma$の範囲に落ちることが知られている。さらに下位の１パーセンタイルの位置は平均からの距離が-2.33σとなり、33パーセン

タイルの位置は－0.44σ、50パーセンタイルの位置は0となり、67パーセンタイルの位置は＋0.44σ、99パーセンタイルの位置は＋2.70σとなる。このような正規分布に関する知識を利用し、相対的な順位を量的な値で特徴づけることができる。つまり、標準偏差σを１単位としてパーセンタイルのような相対的順位を平均からの距離で測ることができる。具体的には次のようにする。まず、賛否程度に従って被験者全員をランクづけして被験者のパーセンタイルを得る。あるパーセンタイルの被験者が標準正規分布の母集団内で、どのような位置にいるかを分布のベース値で特徴づける。標準正規分布であるからZ値と呼んだ方が理解されやすいかも知れない。そのベース値を被験者の点数とする。したがって33パーセンタイルの被験者には－0.44の点数、50パーセンタイルは０の点数、67パーセンタイルは＋0.44の点数をつける。このようにして相対的順位の情報から量的な情報を得る。このような点数化を反応カテゴリー（回答選択肢）に従って順序づけられた被験者グループに対して行う。

そのような手法がリッカートのシグマ法（sigma method）である。まず、各意見項目につき、調査結果から反応カテゴリーの相対度数を得る。そして相対度数に従って標準正規分布を分割する。当然、順位に応じた位置関係にして分割する。したがって分割されたエリアは相対度数に応じた大きさになっており、エリアの位置は順位と相対度数の大きさの関係で決まってくる。分割されたエリアの平均値を、その反応カテゴリーを選択した被験者の評価点（シグマ値）とする。図１はリッカート（1932）が例示している意見項目に関し、反応カテゴリーの相対度数に基づいて標準正規分布を分割したものである。

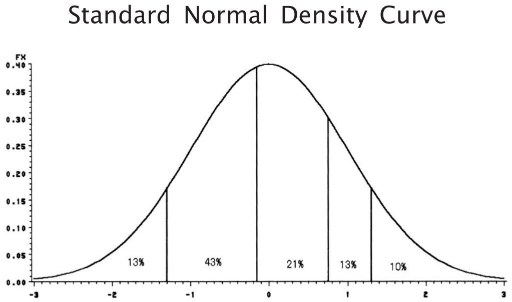

図１　反応カテゴリーの相対度数に従って標準正規分布を分割した図

分割されたエリアの平均値から表１に示すようなシグマ法の値（シグマ値）を得ることができる。リッカートはThorndike（1919）の表22を用いてシグマ値を得ており、それは各エリアが１パーセントになるように標準正規分布を100個に分割し、－3から＋3の範囲をもつ「離散型の標準正規分布」のようにした分布表である。したがって統計学の教科書などに記載されている標準正規分布表に基づくシグマ値とThorndikeの表22に基づくシグマ値は若干異なる可能性があるが、分割されたエリアの平均値をシグマ値とする考え方は同じである。

表１　リッカートのシグマ法と簡便法との関係を示す数値例

	強く賛成	賛　成	どちらでもない	反　対	強く反対
相 対 度 数	13％	43％	21％	13％	10％
シグマ法の値	－1.63	－0.43	＋0.43	＋0.99	＋1.76
原点を移動させた場合の値	1.00	2.20	3.06	3.62	4.39
簡便法の値	1	2	3	4	5

ところで、被験者の散らばりを測る尺度は標準正規分布に限る必要はなく、それを左右に移動させたような正規分布もよい。つまり、平均の位置はゼロに限る必要はなく、別の点でもよい。例えば一番小さいシグマ値－1.63が1.00となるような点数化も考えられる。その場合の配点は表１において原点を移動させた場合の値として示している。その配点を見ると、表１の最終行に示すような整数の配点にして単純化することが考えられる。このようにしてシグマ法の点数化を簡略化したのがリッカートの簡便法である。

リッカートが提案するような簡便法の妥当性として次のような点を示している。まず、簡便法とシグマ法とを比較し、ほぼ同じ程度の信頼係数（偶数項目と奇数項目との間での回答の一貫性）が得られること、両者のデータがほぼ完全に相関していることを示している。具体的にはシグマ法による点数化、1, 2, 3, 4, 5の配点からなる点数化、1, 3, 4, 5, 7の配点からなる点数化を用い、これら三つの点数化から得られたデータの相関係数を報告している。国際主義に対する態度ではデータの相関係数がそれぞれ0.99程度となり、ほぼ完全に相関していることを報告している。異なる被験者グループ、つまり異なる標本を用いても同様な結果になることを示している。黒人に対する態度についても、同様な相関係数を得ている。簡単な点数化を用いるリッカートの尺度法とサーストンの尺度法とを比較し、同程度の信頼係数がえられること、両者のデータが強く相関していることなども示している。

リッカートが開発した尺度構成法では、複数の意見項目に関する被験者のスコアを合計したり平均化したりする。その場合、ある意見項目では「強く賛成」の方がポジティブな態度を意味し、別の意見項目では「強い反対」の方がポジティブな態度を意味することもあるので、「強い反対」と「強い賛成」のどちら側を点数の小さい方にするかは、ポジティブな態度に関する意見項目なのか、ネガティブな態度に関する意見項目なのかによって決まってくる。さらに個々の意見項目に関する被験者の散らばりが正規分布に従うことが仮定されているので、全ての意見項目の点数を線型変換して得る被験者の最終スコアも正規分布に従う。これは正規分布の再生性で保証される。リッカートの例では、国際主義に対する態度尺度、黒人に対する態度尺度、帝国主義に対する態度尺度が、それぞれ正規分布に従うことを意味する。先に示した引用文の最後の文節「use this assumption as the basis for combining the different statements」は、このことを意味している。正規分布からのデータであるから、小標本の場合でもｔ検定やF検定など、正規分布に基づく統計的推論が可能になってくる。

３．リットカート尺度に関する用語の混同

これまでの説明を踏まえ、リッカート尺度を定式化してみよう。測定しようとしている量的特性をWと記し、その測定に用いる意見項目はm個とする。ある意見項目（statement）sに関する被験者iの点数はX_{is}とする。量的特性Wを測るのに採用した全ての意見項目について被験者の点数X_{is}を加算し、その加算点または平均点を当該量的特性Wに関する被験者のスコアとする。したがって量的特性Wに関する被験者iのスコアX_iは次のようになる。

$$X_i = \frac{\sum_{s=1}^{m} X_{is}}{m} \quad \text{または} \quad X_i = \sum_{s=1}^{m} X_{is} \qquad (1)$$

これが、リッカートが意図していた尺度である。この尺度から得るデータは$\{X_i\}_{i=1}^{n}$と記す。ただし、nはデータの大きさ（被験者の数）である。その一方で、単独の意見項目sから作るデータ$\{X_{is}\}_{i=1}^{n}$の場合もリッカート・データやリッカート尺度と呼ばれている。しかし、ここには大きな違いがある。リッカートが意図したところの前者のデータ$\{X_i\}_{i=1}^{n}$は様々に異なった数値からなり、正規分布に従う連続型変数Wのデータに相応しいものになっている。ところが後者のデータ$\{X_{is}\}_{i=1}^{n}$は回答選択肢の数だけの変動しかない。リッカートの例では５通りの変動しか見られない。このような大きな違いあるので、本稿の冒頭で述べたように、単独の意見項目sから作られる後者のデータ$\{X_{is}\}_{i=1}^{n}$は便宜上「リッカート型項目データ」とよぶ。

４．リッカート型項目データの間隔データとしての使用

リッカート型項目データは、数値の間隔がどこでも同等であることが保証されていないから、順位データとして取り扱うべきであると結論づけることは容易である。しかし、多くの研究者がリッカート型項目データを間隔データとして用いて分析を行い、そのことが今後も続くこと

を考えると、本稿の目的で述べたように、リッカート型項目データを間隔データとして用いることが正当化できる条件をわかりやすい形で示すことには意味があると考える。

リッカート（1932）は相対的な順位の情報から量的な情報を作り出すメカニズムとして正規分布を利用している。したがって正規分布に従う母集団が存在することが肝心かなめの条件である。これは単独の意見項目から作るリッカート型項目データの場合でも同様である。したがってリッカート型項目データを間隔データの如くに取り扱うことが正当化できる場合の条件は次のようになる。

リッカート型項目データの散らばりが正規分布の形を反映していると推測できること。換言すると、正規母集団からの標本をグループ化したようなベル型の離散型分布をもったデータであること。

このような条件が満たされているか否かをチェックする必要がある。その一つ方法は適合度のカイ２乗検定である。適合度の検定における一方の分布は現実のデータの分布であり、他方は正規分布である。もっと単純な方法はデータの相対度数分布を示すヒストグラム（柱状グラフ）を描き、正規分布のようなベル型から大きく隔たっていないことを確認することである。

データの散らばりが正規分布の形からある程度かけ離れていても、統計的推論の結果に大きく影響しないことが知られている。このことに鑑み、ここでの判断基準は厳しくする必要はない。しかし、分布の一番端で相対度数が最も大きくなっている場合、データの散らばりが一様分布を反映している場合、ある選択肢に回答が集中している場合など、データの散らばりが正規分布の形を反映していないことが明らかな場合は、正規母集団の仮定は成り立たない。一方、データの散らばりが正規分布の形を反映していると推測できる場合は、多少の誤差が存在するものの、データの散らばりは被験者の散らばりを特徴づけるものとなっている。そのようなリッカート型項目データは間隔データとして取り扱うことに妥当性がでてくる。

データの散らばりが正規分布の形を反映していると推測できることを条件にし、順位データを間隔データとして取り扱うことに妥当性があると考える研究者は他にもいる。それはBorgatta and Bohrnsted（1980）であり、彼らの主張は次のようなものである。

> As it seems to us that most constructs are conceptualized as continuous and can be thought of as reasonably distributed in the population using a bell-shaped curve as a model, we see no reason not to analyze the manifest data using parametric statistic, even though they are imperfect interval-level scales.（Borgatta and Bohrnsted 1980: 160）

これは先に述べたリッカートの論理と同じである。順位の情報から量的な情報を作り出すのにリッカートが正規分布を利用していたことを彼らが知っていたかどうか不明であるが、彼らの考え方はリッカートの論理に沿ったものである。Borgatta and Bohrnstedの考え方に同意し、それを実証的な研究に利用したものとしては、例えばMorrison and Toy（1982）がある。

なお、潜在変数の分布としては正規分布以外の分布も考えられる。順位の情報から量的な情報を得るのに一様分布を利用するものとしてはBross（1958）のRidit法がある。

５．点数化と反応カテゴリーの数に関する考察

1, 2, 3, 4, 5のように等間隔で点数化してリッカート型項目データを作り、その相対度数分布が正規分布の形を反映していると推定できる場合は、等間隔の点数化を利用すればよい。一方、正規分布の形を全く反映していないと言うほどでもないが、かなり正規分布の形から崩れているような場合は、どのような配点の工夫が考えられるのであろうか。

理想的にはシグマ法を用いるのがよい。ある潜在変数Wに対して意見項目が20個も、30個もあれば全ての意見項目に対してシグマ値をそれぞれ算出することは多大の労力を要する。しかし、本稿で問題にしているのは意見項目が一つの場合であるので、シグマ値の算出には多大な労力を要することにはならない。またリッカートの時代と異なり、現代のコンピュータの性能と普及を考えると、シグマ値を算出することは比較的容易である。参考までにシグマ法の計算式を示す（Golden and Brockett 1987）。

$$Z_j^* = \frac{1}{f_j\sqrt{2\pi}}\left\{\exp\left(-\frac{Z_{j-1}^2}{2}\right) - \exp\left(-\frac{Z_j^2}{2}\right)\right\} \qquad (2)$$

ただし、Z_j^*は反応カテゴリーjのシグマ値、f_jは反応カテゴリーjの相対度数、Z_{j-1}とZ_jは標準正規分布において当該反応カテゴリーjが占めるエリアの下限Z値と上限Z値である。このようにコンピュータを利用すると、シグマ値を算出することは比較的容易である。

とは言うものの、シグマ値を算出しないで済むようなリッカート式の簡便な点数化を考えてみよう。データの相対度数分布を示すヒストグラム（柱状グラフ）と正規分布を示すベル型のグラフとを重ね合わせた状態を想像してほしい。分布の端に近いところで、柱の高さが正規分布の高さに比べて極端に高くなっている場合は、その柱の分布端側に２倍の間隔を配置することが考えられる。つまり、正規分布に比べて極端に高くなっている柱は押しつぶし、その柱の幅を広げる考え方である。別の表現にすると、２倍の間隔をとったところはデータの密度を２倍に薄める考え方である。このような工夫はシグマ法の考え方に沿ったものである。ランキングされた被験者の散らばりを正規母集団の形に当てはめて被験者の点数とするのがシグマ法の考え方であるから、正規分布に比べて極端に高くなっている部分は押しつぶしてならす考え方である。

一般的に考え、幾らかの意見項目を合成して作るリッカート・データの場合は、左や右に歪んだ分布をもつ意見項目が存在しても、それらの線型組み合わせで出来ているので、歪みが小さくなることがある。しかし、単独の意見項目から作るリッカート型項目データの場合には、そのようなメカニズムが存在しない。したがってリッカート型項目データが「正規性」を満たすケースは相対的に少ないことが想像できる。リッカート型項目データの分布が歪みを持ったものになる傾向があることは、例えばRowan（2001）で指摘しされている。具体的には人々はポジティブ（賛成）側の回答選択肢、しかし一番端でない回答選択肢を選ぶ傾向がある。このような場合は既に述べたような点数化の工夫を試みることが考えられる。

次は反応カテゴリーの数について考えてみよう。リッカート型項目データをつくるにあたり、反応カテゴリーの数は中性点のない４個や６個でもよいか否かが議論されている。上に述べたような論理からすると、一般論として反応カテゴリーの数は奇数でも偶数でもよい。研究の具体的な内容によって決ってくる問題である。「どちらでもない」といった中性点を選択している場合でも、「判断できない」といった意味をもつ内容の場合もある。このような中性点に関する研究としては、例えばRaaijmakers（2000）がある。

さらに反応カテゴリーの数はどの程度が適切であるかも議論されている。反応カテゴリーの数は３個でも良いのかといった疑問がその例である。リッカート型項目データを間隔データとして使う以前に、そのデータが正規母集団から抽出されたものであるとの検証が必要である。反応カテゴリー数が３個といったように少ない場合は、そのような検証が困難になる。さらに正規分布に従う連続型の賛否程度を３段階の賛否程度に集約することは、やり過ぎといった感がある。したがって反応カテゴリーの数は少なくとも５個になるようにすることが考えられる。正規母集団であることを考えると、反応カテゴ

リーの数は多いほどよいが、その一方で反応カテゴリーの数が多くなりすぎると、被験者が反応カテゴリーを選択するときの信頼性が落ちてくる。したがって反応カテゴリーの数は一般的には５個ないしは７個程度がよいと思われる。

反応カテゴリーの数はどの程度が最適であるかといった研究としては、例えばBirkett（1986）がある。彼は反応カテゴリーが２個、６個、14個の場合について比較し、反応カテゴリーが６個の場合に回答の信頼性が高くなる傾向があるとする経験的な結果を示している。

６．リットカート型項目データの誤差に関する考察

リッカート型項目データを間隔データとして用いる場合、データは誤差を含んだものとなる。そのような誤差がどのようなものであるかを考察する。これは後のセクションで議論することの準備でもある。

ここでは単独の意見項目から作り出されるデータが興味の対象であるので、記号$\{X_{is}\}_{i=1}^{n}$に代え、意見項目の記号sを省略した記号$\{X_i\}_{i=1}^{n}$を用いる。リッカート型項目データ$\{X_i\}_{i=1}^{n}$は次のように分解することができる。

$$X_i = W_i + u_i \quad \text{for} \quad i = 1, 2, \ldots, n \tag{3}$$

ただしW_iは被験者iの賛否の程度を真に特徴づける正規変数Wの値であり、u_iは誤差項である。標本の採り方は無作為抽出に従うものとする。したがってある被験者iの点数X_iと別の被験者hの点数X_hとは独立している。さらにとW_iとu_iは独立しているものとする。

ここでの主たる興味はシグマ法の配点から簡便法の配点にした場合に、どのような誤差が生まれるかにある。したがってシグマ法が基準であるとの観点に立ち、基準の尺度値からの逸脱がどのような結果をもたらすかに焦点をおく。シグマ法は被験者の順位情報を正規母集団の形に当てはめて点数をつけることから明らかなように、シグマ法に基づく被験者の平均は正規母集団の平均$E(W)$と等しいものとして設計されている。

上に示したリッカートの例を用い、一般性を失うことなく議論を進める。シグマ法の配点を、原点を変えて表現したものは次のようになっている。

1.00，2.20，3.06，3.62，4.39

便宜上、これをシグマ法の結果とする。点数間隔が不均一であるが、この配点には被験者たちの賛否の程度を特徴づける重要な情報が含まれている。このようなシグマ法の点数はW_j^*（$j=1, 2, \ldots, c$）と記す。これに対して次のような一定間隔の配点を用いたとしよう。

1，2，3，4，5

このような簡便法の配点はX_j（$j=1, 2, \ldots, c$）と記す。

ここでは誤差項u_iを２つのタイプの誤差に分解する。第１のタイプの誤差は例えば正規変数Wの値4.20を「強く反対」の平均値4.39にすることから生じる誤差である。つまり、正規変数Wの様々な実数値W_iをグループ化してc個のカテゴリー代表値W_j^*（$j=1, 2, \ldots, c$）にすることから生じる誤差である。ここでj番目の反応カテゴリーの相対度数をf_jと記すと、シグマ法の平均は

$$\overline{W}^* = \sum_{j=1}^{c} f_j \cdot W_j^* \tag{4}$$

で与えられる。だたし $\sum_{j=1}^{c} f_j = 1$ である。一方、賛否の程度を真に特徴づける値W_iは実際には観測できないが、その平均は

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i \tag{5}$$

で与えられる。反応カテゴリーの代表値が平均値であるから、$\overline{W}^* = \overline{W}$となるようにシグマ法が設計されている。このことに鑑み、第１のタイ

プの誤差は平均でゼロとする。

第２のタイプの誤差は例えば「強く反対」の代表値4.39を整数５にすることから生じる誤差である。つまり、シグマ法の配点を簡便法の配点にすることから生じる誤差、いわゆる尺度の歪みに起因する誤差である。この種の誤差は被験者全員に対して共通した誤差である。リッカートの例では４箇所で尺度歪みが生じ、

0.00, －0.20, －0.06, 0.38, 0.61

となっている。このような尺度歪み（点数間隔の違い）を$\delta_j = X_j - W_j^*$ $(j=1, 2, \ldots c)$と記す。このような尺度歪みは図２で特徴づけることができる。シグマ法の配点（1.00, 2.20, 3.06, 3.62, 4.49）を横軸、簡便法の配点（1, 2, 3, 4, 5）を縦軸にとり、ふたつの配点の関係を示している。図２に示す関係が直線であれば尺度歪みが全くない状態である。直線でないことからリッカートの簡便法の配点（1, 2, 3, 4, 5）にはある程度の尺度歪みがあることが視覚的に理解できる。

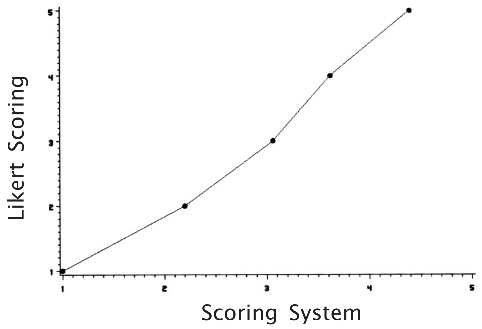

図２　リットカートのシグマ法と簡便法との関係を特徴づける図

尺度歪みによる誤差が点数の平均化によってどのようになるかを調べて見よう。リッカートの例において相対度数を用いて平均を算出すると、シグマ法の平均は$\overline{W}^* = 2.63$となるのに対して簡便法の平均は$\overline{X} = 2.64$となる。したがって誤差の平均$\bar{u}$は0.01となる。上に示すように若干大きな尺度歪みがあると思われる場合でも、誤差の平均$\bar{u}$が0.01と小さくなるのは、平均をとる過程で誤差のプラス・マイナスが相殺されるためである。したがって誤差の平均$\bar{u}$がどのような構造で成り立っているかを分析する必要がある。簡便法の平均$\overline{X}$とシグマ法の平均$\overline{W}^*$とを対比させると次のようになる。

$$\overline{X} = \sum_{j=1}^{c} f_j \cdot X_j \qquad \text{対} \qquad \overline{W}^* = \sum_{j=1}^{c} f_j \cdot W_j^* \tag{6}$$

ここで議論しているタイプの尺度構成法においては線型変換によって異なったスケールで尺度値を表現できる。そのため先の例では「強い賛成」において簡便法の点数とシグマ法の点数とを等しくしている。したがって一般化した議論においても$X_1 = W_1^*$と設定する。その結果、誤差の平均は次のように分解できる。

$$\begin{aligned} \bar{u} &= \overline{X} - \overline{W} = \overline{X} - \overline{W}^* = \sum_{j=1}^{c} f_j \cdot X_j - \sum_{j=1}^{c} f_j \cdot W_j^* \\ &= \sum_{j=1}^{c} f_j (X_j - W_j^*) = \sum_{j=2}^{c} f_j \cdot \delta_j \end{aligned} \tag{7}$$

したがって誤差の平均$\bar{u}$は、先に述べた尺度歪みδ_jと相対度数f_jによって決まってくる。なお、$X_1 = W_1^*$であるから$f_1 \delta_1 = 0$となっている。

誤差の平均$\bar{u}$がゼロとなるのは次のような場合である。その第１のケースは簡便法の点数間隔がシグマ法の点数間隔と全ての場所で偶然的に一致し、したがって尺度歪みが全ての場所でゼロ$(\delta_j = 0)$になっている場合である。第２のケースは尺度歪みがプラスとマイナスとに分かれており、相対度数との組み合わせで、うまくプラス・マイナスが相殺されて$\bar{u} = 0$となる場合である。正規分布の形と比較してデータの相対度数分布が左右対称な形で崩れている場合が第２のケースの例となる。

リッカート（1932）が用いた例は第２のケースに近い。リッカートの例ではプラスとマイナスの尺度歪みが半々の割合で生じており、平均を算出するプロセスで誤差のプラス・マイナスがかなりの割合で相殺されるようになっている。尺度歪みだけのプラス・マイナスの相殺は＋0.73となるが、マイナスの尺度歪みに対応する相対

度数が大きく、プラスの尺度歪みに対応する相対度数が小さいことから、尺度歪みだけで見る場合よりも、実際の誤差のプラス・マイナスの相殺は大規模で生じている。その結果、誤差の平均$\bar{u}$が0.01と小さくなっていることが説明できる。

データの相対度数分布が左や右に大きく歪んでくると、誤差のプラス・マイナスの相殺がアンバランスな形で生じ、誤差の平均$\bar{u}$はゼロから離れる方向へ変化する。これは正規母集団からの標本とは思えないほどに大きく歪んだデータを間隔データとして用いると、どのような結果になるかを示唆している。

7．Labovitz式の点数化に関する考察

ここでは順位に従って任意に点数をつけた順位データをほぼ無条件で間隔データの如くに取り扱ってもよいとする議論について考察する。それはLabovitz（1967, 1968, 1970, 1971, 1972, 1975）の議論である。良く引用されているのは彼の1967年と1970年の文献である。Labovitz（1967）は次のように述べている。

> Treating ordinal data (which may or may not be approximately interval) as interval data by arbitrarily assigning numbers to the ordinal categories can be both legitimate and useful.（Labovitz 1967: 153）

当然のことながら、点数化は順位に従って点数が単調に増えるか減るかの単調な関係になっている必要がある。特に順位の違いの大きさに関する知識が無い場合や少ない場合は、点数間隔が一定になっている線型関係が最もよい選択であるとしている。1970年の文献では、点数化はシステマッチクに行っても、ランダムに行っても良く、順位データを間隔データの如くに取り扱うことができるとしている。

このような結論を鵜呑みにし、順位に従って任意に点数をつけた順位データを無条件で間隔データとして取り扱っても良いと解釈しない方がよい。彼の結論には幾らかのただし書きが必要である。

第１のただし書きは、点数化の違い（配点の違い）が消えてなくなるような加工をした後の集計情報に基づいて結論を得る場合は、彼のような結論を得ることが可能である。彼の結論は点数化の違い（配点の違い）がピアソンの相関係数に大きく影響しないことが第１の根拠になっている。ピアソンの相関係数は一方の変数が増えたときに別の変数が増えるか減るかと言った連動性、具体的には線型の連動性を測るものである。さらにピアソンの相関係数は、考慮している変数のスケールに関わりなく、－1から＋1の範囲の値になるような相対尺度のメカニズムをもっている。変数の数値が平均よりも大きいか小さいかと言ったプラス・マイナスの符号、二つの変数間でのプラス・マイナスの符合の積、その積の頻度などが重要な情報である。これは点数のスケールの違いなどは重要な情報でないことを意味する。重要な情報は汲み取り、そうでない情報は消し去るようなメカニズムになっている。このような相関係数のメカニズムのため、点数のスケールの違いは完全に修正されるというか完全になくなる。一方、異なる点数化による配点の違いの方は完全には消されないかもしれないが、先に述べたような単調関数の関係になっている場合はかなりの程度が消されていく。そうであるからこそ、配点の違いが相関係数に大きく影響しないのである。もうひとつの根拠は、二つの処置法に対する評価の平均の差が有意であるかどうかの検定において、点数化の違い（配点の違い）が検定の結果に大きく影響しないことを挙げている。ここでは平均を算出して平均の差を取り、それを標準化したものが検定統計量となる。このような検定統計量も相対的な尺度としてのメカニズムをもっている。標準化することで、点数のスケールの違いが結果に影響することはない。例えば（1, 2, 3, 4, 5）に代えて点数間隔を２倍にした（1, 3, 5, 7, 9）の配点を用いても、同じ結果が得られる。一方、異なった点数化による配点の違いの方は、平均

を算出して差をとる過程でかなりの程度が消えてゆく。先に示した尺度歪みとは配点の違いのことであり、平均を算出する過程で誤差のプラス・マイナスが相殺されることは既に示した。これは配点の違いが消されていくことと同じである。そして標準化を終えた状態では、配点の違いは結果に大きく影響しないようになっている。ここでは二つの処置法に対する評価の平均が異なっているかどうかが重要な情報であり、重要な情報は汲み取り、そうでない情報は消し去るようなメカニズムになっている。そうであるからこそ、配点の違いが検定結果に大きく影響しないのである。別の言い方をすると、平均や標準偏差など、点数化の違いが消えていない状態の統計量に基づく分析では、点数化の違いが結論に大きく影響する。Labovitz（1967）が示す例においても、点数化の違いが平均に大きな差を与えている。

第２のただし書きは、被験者の散らばりを真に特徴づける配点と実際に用いる配点との関係が線型関係から極端に崩れてL字状のような関係になっていないことである。例えば一方の配点がほぼ一定間隔であるのに対し、他方の配点では一番端側の間隔が極端に大きく他の間隔が極端に小さくなっている場合は、L字状のような関係になる。われわれは順位の違いの大きさに関する知識を全く持っていない場合や少ししかもっていない場合が普通である。そのような状況では様々な間隔の点数化を用い、点数化の違いが結果に大きく影響するかをチェックすることになる。Labovitzの手法もこれと同じであり、実際に用いた様々な点数化の違いが相関係数などの結果に強い影響を与えるかどうかをチェックしているに過ぎない。Labovitz（1967）の第２表に示す７つの点数化において、等間隔からなる第７番目の点数化の配点$(0, 3\frac{1}{3}, 6\frac{2}{3}, 10)$を横軸にし、他の６個の点数化の配点を縦軸にし、配点相互間の関係を見てみると図３のようになる。

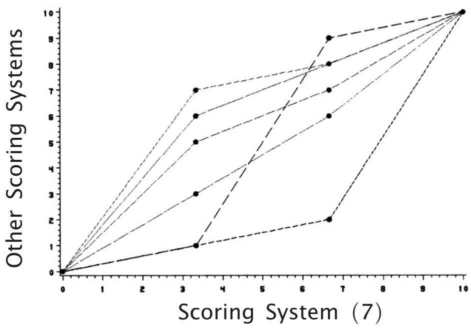

図３　Labovitzka（1967）の点数化を特徴づける図

図３に示されている関係はいずれも線型ではないが、データを直線であてはめる観点（相関的な観点ないしは回帰的な観点）から見ると、かなり線型に近い関係になっている。直線から最も崩れた関係でも、データ点を直線であてはめる観点から見ると、かなり強い相関があることが容易に想像できる。事実、線型に近い関係になっているペアーほど、Labovitz（1967）の第４表に示されている相関係数が大きくなっている。第４表の相関係数は実際に用いた７つの点数化の配点に強い相関があることを示しているに過ぎない。また、Labovitz（1970）は点数化をシステマッチクに行っても、ランダムに行っても良く、順位データを間隔データの如くに取り扱うことができるとする結論は変わらないとしている。ランダムの点数化の場合でも、図３に示すような単調な関係が配点間に維持されておれば、お互いに相関の強いものになることは容易に想像できる。しかし、それは実際に使用した様々な点数化の配点に強い相関があることを示しているに過ぎず、被験者の散らばりを真に特徴づける配点と実際に用いる配点との関係は不明のままである。不明である以上、実際に用いた様々な点数化の配点間で強い相関があることを根拠にして、被験者の散らばりを真に特徴づける配点と実際に用いる配点との間に強い相関があるとは言えない。さらに線型関係から大きく隔たりL字状のような２極化（dichotomy）の関係になってくると相関が弱くなり、Labovitz（1970, 1971）が認めているように彼の結論は成り立たなくなる。このことも、図３から容易に理解できる。

被験者の散らばりを真に特徴づける配点が図３の横軸の配点であると仮定すると、縦軸に示す点数化の配点には大きな尺度歪みを持ったものが存在している。直線から隔たったものほど、大きな尺度歪みを持っていることを意味する。このような大きな尺度歪みがあったとしても、平均化などのプロセスを経ると、測定誤差のプラス・マイナスが相殺されることを利用したのがLabovitzの論理である。

Labovitzの結論には少なくとも上に述べたような前提条件が必要である。Labovitzの結論には隠された前提条件があることは他の研究者も指摘している。例えばO'Brien（1979）はLabovitz（1970）が実験に使った条件とは異なる条件でもLabovitzのような結論が得られるかを実験した。彼が見つけた結果のひとつは、ランダムな点数化に用いる分布によって相関係数が大きく異なり、正規分布と一様分布の下では大きかった相関関係が、歪んだ分布の下では大幅に小さくなる。これは一例にすぎない。Labovitzの結論には幾らかの隠された前提条件がある。したがってLabovitzを引用するだけでもって、直ちに順位データを間隔データとして取り扱っても良いという訳にはいかない。

８．Likert式点数化とLabovitz式点数化との比較

ここでは意見項目に関するリッカート式の点数化（簡便法）とLabovitz式の点数化とを比較し、その類似点や相違点について考察する。

被験者の散らばりを真に特徴づける配点と実際に用いる配点との関係が線型関係から大きく逸脱していなければ、順位データを間隔データの如くに取り扱うことの正当性を得ることができる。これはリッカート式の点数化とLabovitz式の点数化との類似点である。

両者の重要な第１の相違点は、被験者の散らばりを真に特徴づける配点と実際に用いる配点との関係が不明であるか否かに関わる。Labovitz式の点数化では、そのような関係が不明のままであるので、線型関係から大きく逸脱しているのか否かをチェックすることができない。そのため実際に用いる配点が妥当なものであるかの保障はされておらず、順位データを間隔データとして用いることの正当化に関して不安が残る。これに対してリッカート式の点数化は正規母集団の仮定を用いることから、データの散らばりが正規分布の形を反映していると推測できる場合は、被験者の散らばりを真に特徴づける配点と実際に用いる配点との関係が線型関係から大きく逸脱していないことを意味し、実際に用いる配点が妥当なものであるとの保証が得られる。

つまり、Labovitz式の点数化では尺度の歪みの大きさが不明のままであるが、正規母集団という追加の情報を加えることで、そのようなことは不明でなくなる。データの散らばりが正規分布の形を反映していると推測できる場合は、尺度の歪みはある程度小さい範囲に収まっていることが保証されている。正規母集団という情報を加えることで、順位データを間隔データとして用いることができるか否かの実用的な判断基準を与えてくれる。

重要な第２の相違点は尺度歪みの大きさに対する姿勢の違いである。被験者の散らばりを真に特徴づける配点と実際に用いる配点との比較において、配点の違いは「尺度歪み」でもある。図２と図３において、被験者の散らばりを真に特徴づける配点が横軸の配点であると仮定すると、直線から大きく崩れているほど、縦軸の配点には大きな尺度歪みが起こっていることを意味する。図２と図３との比較から、尺度歪みに対する両者の姿勢の違いが読み取れる。Labovitz式の点数化は大きな尺度歪みをも許すものである。大きな尺度歪みが様々な場所で存在しても、相関係数や平均の差の検定統計量など、集計された結果に大きな影響を与えない可能性があり、そのような可能性は極めて大きく、したがって任意に点数化したデータを間隔データとして使っても良いと言うのがLabovitzの結論である。しかし、Labovitz式の点数化によって得た個々の被験

者の点数は、大きく歪んでいる可能性が高く、信頼できるものではない。このことは彼が使っている具体例で見ると、その姿が良く見えてくる。ある処置に対する患者の評価を０点から10点の範囲で測る場合、同じ患者の回答が、ある点数化では10点満点中２点となり、別の点数化では10点満点中９点となる。同じ患者の評価が10点満点中２点でも９点でも良い。これがLabovitz式の点数化である。このようなものが、ものを測る尺度と呼べるか疑問である。

歪みのない尺度を使うのが理想的であるが、そのような尺度を得ることが不可能か困難であるならば、次善の策として歪みがどのようなところでも小さい尺度を使うべきである。そうでなければ尺度とは言わない。リッカート式の点数化においては、データの散らばりが正規分布の形を反映していると推測できる場合は、尺度歪みがある程度小さい範囲に収まっている保証がある。したがって、個々の被験者の評価点はLabovitz式の点数と比較して高い信頼性がおける。

９．おわりに

リッカートは相対的な順位の情報を量的な情報へ変換するメカニズムとして正規分布を使っている。それは正規分布に従う被験者の散らばりをグループ化して整数の配点で特徴づける形で利用されている。したがって正規母集団の仮定がリッカート尺度構成法のかなめである。正規母集団の仮定については様々な議論があるので、正規母集団を仮定できないような場合は無視しても良いというようなものではない。リッカート型項目データを間隔データとして取り扱うのであれば、データの散らばりが正規分布の形を反映していると推測できることが必要である。そのような検討をしないまま間隔データとして用いることは、「リッカート尺度もどき」のものを作ってリッカート尺度として使うことに等しい。さらに小標本の場合はｔ検定やF検定など、正規分布に基づく統計的推論を行うのが一般的であることを考えれば、別の意味でも正規母集団の仮定は重要である。

リッカート尺度を紹介している教科書的な本を幾らか調べた（Mann 1968, 宝月ほか1989,安田・原1982, 福武1984）。リッカート尺度をつくる手順を説明しているが、どのようにして順位の情報から量的な情報を作り出すのかといった肝心な点が論理的に説明されているとは言いがたい。したがってリッカート式の点数化によって得たデータがどのような条件のときに間隔データとして利用できるかが不明である。例えば福武（1984）ではシグマ法を説明するところで正規母集団の仮定のことを言っているが、簡便法のところでは任意に1, 2, 3, 4, 5といった点数を与えるとなっている。したがって正規母集団の役割が不明である。このような状態であるからこそ、順位の情報を量的な情報へ変換するメカニズムとしての正規母集団の役割を強調するものである。

つぎはリッカート型項目データを間隔データとして用いる場合、どのような誤差が含まれるかを考察した。そして正規母集団の仮定の重要視する観点からLabovitz（1967, 1970）の点数化に関し、批判的な分析を試みた。彼の点数化では実際に用いる配点が妥当なものであるかの保証はされておらず、順位データを間隔データの如くに取り扱うことの正当化に関して不安が残る。さらに個々の被験者の点数は全く信頼の置けないものになる。このような尺度としての重大な欠陥は、正規母集団の仮定を用いることで修正できる。特にリッカート式の点数化から得たデータを間隔データの如くに取り扱うことができるか否かの実用的な判断基準を与えてくれることは重要である。

（青森公立大学）

（2004年11月９日受付、2004年12月２日受理）

参考文献

Abelson, Robert P., and John W. Tukey (1970), "Efficient Conversion of Non-Metric Information into Metric Information," *The Quantitative Analysis of Social Problem* (ed. Edward R. Tufte), Addison-Wesley, 407-417.

Agresti, Alan (1990), *Categorical Data Analysis*, John Wiley & Sons.

__________ (1996), *An Introduction to Categorical Data Analysis*, John Wiley & Sons.

Birkett, Nicholas J. (1986), "Selecting the Number of Response Categories for a Likert-type Scale," *Proceeding of the Survey Research Methods Section, American Statistical Association*, 488-492.

Borgatta, Edgar F., and Geoge W. Bohrnstedt (1980), "Level of Measurement: Once Over Again," *Sociological Methods and Research* 9 (November), 147-160.

Brown, James Dean (2000), "What Issues Affect Likert-scale Questionnaire Formats?" (http://www.jalt.org/test/bro_7.htm: July 2003).

Bross, Irwin D. J. (1958), "How to Use Ridit Analysis," *Biometrics* 14 (March), 18-38.

Champion, Dean J. (1968), "‘Some Observations on Measurement and Statistics’: Comment," *Social Forces* 46 (June), 541.

Chase, Clinton I. (1978), *Measurement for Educational Evaluation* (Second Edition), Addison-Wesley.

Clason, Dennis L., and Thomas J. Dormody (1994), "Analyzing Data Measured in Individual Likert-Type Items," *Journal of Agricultural Education* 34 (4), 31-35.

Downie, N. M. (1967), *Fundamentals of Measurement: Techniques and Practices*, Oxford University Press.

Easthope, Gary (1974), *A History of Social Research Methods*, Longman. (=1982,阿久津昌三, 他４名訳『社会調査方法史』慶應通信).

Fielding, Antony (1997), "On Scoring Ordered Classifications," *British Journal of Mathematical and Statistical Psychology* 56, 285-307.

Golden, Linda L. and Patrick L. Brockett (1987), "The Effect of Alternative Scoring Methods on the Analysis of Rank Order Categorical Data," *Journal of Mathematical Sociology* 12 (4), 383-414.

Henkel, Ramon E. (1975), "Part-whole Correlations and the Treatment of Ordinal and Quasi-interval Data as Interval Data," *Pacific Sociological Review* 18 (1), 3-26.

Jacoby, William G. (1999), "Levels of Measurement and Political Research:An Optimistic View," *American Journal of Political Science* 43 (1), 271-301.

Kim, Jae-On (1975), "Multivariate Analysis of Ordinal Variables," *American Journal of Sociology* 81 (2), 261-298.

Labovitz, Sanford (1967), "Some Observations on Measurement and Statistics," *Social Forces* 46 (December), 151-160.

__________ (1968), "Reply to Champion and Morris," *Social Forces* 46 (June), 543-544.

__________ (1970), "The Assignment of Numbers to Rank Ordered Categories," *American Sociological Review* 35(June), 515-524.

__________ (1971), "In Defense of Assigning Numbers to Ranks," *American Sociological Review* 36 (June), 521-522.

__________ (1972), "Statistical Usage in *Sociology: Sacred Cows and Rituals," Sociological Methods and Research* 1 (August), 13-37.

__________ (1975), "Comments on Henkel's Paper: The Interplay Between Measurement and Statistics," Pacific Sociological Review 18 (January), 27-35.

Likert, Rensis (1932), *A Technique for the Measurement of Attitudes*, New York: Archives of Psychology.

Mann, Peter H. (1968), *Methods of Sociological Enquiry*, Basil Blackwell. (=1982, 中野正大訳『社会調査を学ぶ人のために 』世界思想社).

Mayer, Lawrence S. (1971), "A Note on Treating Ordinal Data as Interval Data," *American Sociological Review* 36 (June), 519-520.

Morris, Raymond N. (1968), "‘Some Observations on Measurement and Statistics’: Further Comment," *Social Forces* 46 (June), 541-542.

Morrison, Donald G., and Norman E. Toy (1982), "The Effect of Grouping Continuous Variables on Correlation Coefficients," *Marketing Science* 1 (4), 379-389.

O'Brien, Robert M. (1979), "The Use of Pearson's R with Ordinal Data," *American Sociological Review* 44 (October), 851-857.

Raaijmakers, van Hoof, Hart, Verbogt, and Vollebergh (2000), "Adolescents' Midpoint Responses on Likert-Type Scale Items: Neutral or Missing Values?" *International Journal of Public Opinion Research* 12 (2), 208-216.

Roberts, James S., James E. Laughlin, and Douglas H. Wedell (1999), "Validity Issues in the Likert and Thurstone Approaches to Attitude Measurement," *Education and Psychological Measurement* 59 (2), 211-233.

Rowan, John (2001), "Heavier Than Ultramarine?" *The Psychologist* 14(12), 624.

Thorndike, Edward L. (1919), *An Introduction to the Theory* *of Mental and Social Measurements* (Second Edition), Columbia University Press.

Thurstone, L.L., and E. J. Chave (1929), *The Measurement of Attitude*, University Chicago Press.

Schulz, E. Matthew, and Anji Sun (2001), "Controlling for Rater Effect When Comparing Survey Items With Incomplete Likert Data," *CT Research Report Series 2001-2*, (February).

Schweitzer, Sybil, and Donald G. Schweitzer (1971), "Comment of the Pearson r in Random Number and Precise Function Scale Transformation," *American Sociological Review* 36 (June), 518-519.

Vargo, Louis G. (1971), "Comment of 'The Assignment of Numbers to Rank Order Categories'," *American Sociological Review* 36 (June), 517-518.

Wilson, Thomas P. (1971), "*Critique of Ordinal Variables*," Causal Models in the Social Sciences (ed. H. M. Blalock), New York: Aldine, 415-431.

青木繁伸(2000),「多変量分散分析をカテゴリカルなデータで行えるか？」（http://aoki2.si.gunma-u.ac.jp/lecture/mb-arc/arc010/100.html#101:July 2003）.

安田三郎・原順輔（1982),『社会調査ハンドブック　第３版』 有陽閣双書.

林知己夫（1993),『行動計量学序論』 朝倉書房.

福武　直（1984),『社会調査　補訂版』 岩波全書.

宝月　誠・中道 寛・田中 滋・中野 正大（1989),『社会調査』 有陽閣Sシリーズ.

Abstract

Likert (1932) developed a scale for the measurement of attitudes, which became known as the "Likert scale." Nowadays, data extracted from a single Likert-type item are often treated as interval data under the name of Likert scale or Likert data. This is unfortunate because such data are illegitimate Likert data. They are ordinal data by construction. In order to distinguish from legitimate Likert data, data extracted from a single Likert-type item are called "Likert-type item data" in this paper. This paper investigates the logic of the Likert scale in order to find out whether or not Likert-type item data can be treated as interval data under a certain condition. In Likert's book, the normal population is used as a system of transformation from rank information to quantitative information. This mechanism can be used for the case of Likert-type item data as well. From the viewpoint of emphasizing the importance of normal population, this paper shows a practical condition under which Likert-type item data can be treated as interval data. Furthermore, this paper provides answers to a few questions discussed among researchers using Likert-type item data. Labovitz (1967, 1970) demonstrates that treating ordinal data as interval data by arbitrarily assigning numbers to ordinal categories can be both legitimate and useful. This paper critically analyzes the logic behind Labovitz's idea. Such analysis recognizes the importance of the normal population assumption.