リッカート型項目データの回帰への使用と通常最小２乗推定量

村尾　博※

１．はじめに

本稿はリッカート式の点数化から得たデータを量的データとして回帰に用いる場合の通常最小２乗推定量について考察する。

ここで言うリッカート式点数化のデータとは、社会科学の分野で広く一般的に用いられている次のようなアンケート調査のデータである。幾らかの意見項目（言説）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といった具合いに点数化してゆき、「強い賛成」に＋3をつける。別の例では「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。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る。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ここでの議論の対象である。

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

※青森公立大学

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実証的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、単純な記述統計に限らず、相関や回帰などの統計的推論にも見られる。

リッカート式の点数化は点数に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が、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。

リッカート尺度を説明する多くの文献では説明されていないが、リッカート（1932）のオリジナルな文献を読み彼の論理を再構築していくと、どのような条件が満たされれば、リッカート式点数化によって得たデータを間隔データとして用いることができるかが分かってくる。それは、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である[1)]。

ここでは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できるような条件は満たされているとする。そのようなデータであったとしても、回帰分析に用いる場合、通常最小２乗推定量は不偏性を持っているのであろうか。さらに一致性はあるのであろうか。このような疑問が本稿の出発点になっている。もし、通常最小２乗推定量が不偏性を失い、一致性も失うような結果になるのであれば、通常最小２乗推定に基づく実証的研究の結果は信頼の置けないものとなる。したがって本稿ではリッカート型項目データを回帰に用いる場合、特に説明変数側に用いる場合の、通常最小２乗推定量の不偏性と一致性について考察する。そして通常最小２乗推定量が不偏性を失い、一致性も失う場合は、実用的な一つの対処策を提案する。

一方、本稿に関連している研究としてはGolden and Brockett (1987)がある。彼らの研究は順位データに関して幾らかの手法で数量化を行ない、異なる数量化の間で回帰の推定結果を比較している。その中には冒頭で述べたリッカート式の点数化や、後ほど説明するリッカート（1932）のシグマ法も含まれている。彼らの興味はある事例において回帰の係数推定値・分散推定値・決定係数などがどの程度似ているかといったことに置かれており、通常最小２乗推定量の不偏性や一致性が保証されているかといったことに目を向けるものではない。

1）この点に関しては別の論文で考察したので、ここでの更なる議論の展開は控える。

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、リッカート尺度の構成法を説明する。それを受け、セクション３ではリッカート尺度の定式化を試みる。さらにリッカート・データと呼ばれるデータには、性質の異なる２つのタイプがあることを示す。セクション４で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の測定誤差に関する考察を展開する。この考察を受け、セクション５ではリッカート型項目データを回帰に用いる場合の、通常最小２乗推定量の不偏性や一致性について考察する。回帰の説明変数側に用いる場合、それは不偏性を失い、一致性もなくなる結果となる。そこでリッカート型項目データを回帰の説明変数側に用いる場合の、一つの対処案を提案する。本稿のまとめをセクション６で行なう。

２．リッカート尺度の説明

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なう。調査結果に基づき、意見項目ごとに被験者の点数を得る。被験者ごとに全ての意見項目の点数を加算し、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、それを当該量的特性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（百分位）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、67パーセンタイルの位置は$+0.44\sigma$、99パーセンタイルの位置は$+2.70\sigma$となる。このような正規分布の形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（百分位）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パーセンタイルには0の点数、67パーセンタイルの被験者には$+0.44$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なう。

そのような手法の一つ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はリッカートが例として使っている意見項目の相対度数に基づいて標準正規分布を5個に分割したものである。

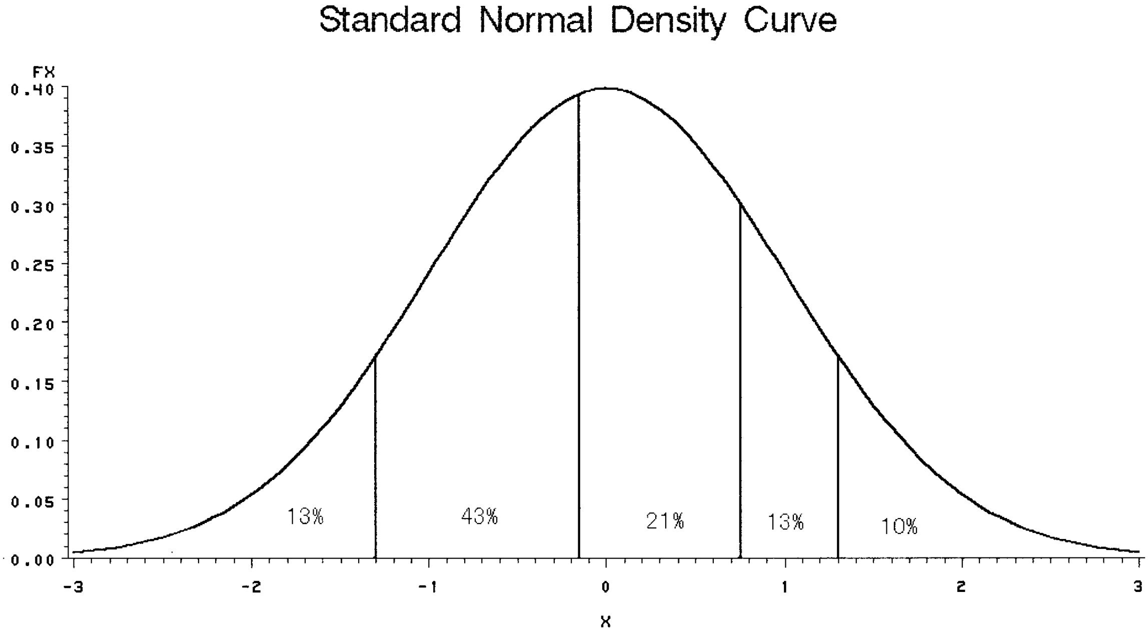

図1．リッカート（1932）のシグマ法

分割されたエリアの平均値から次の表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。

表1．　シグマ法と簡便法

	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

なお、リッカートはThorndike(1919)の表22を用いてシグマ値を得ている。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。

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。その場合の配点は次のようになる。

1.00,　2.20,　3.06,　3.62,　4.39

この配点を見ると、次のような整数の配点にして単純化することが考えられる。

1,　2,　3,　4,　5

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。

リッカートは５個の選択肢を持つ意見項目には(1,2,3,4,5)からなる配点を採用している。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。したがって「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。

リッカートが提案するような簡単な点数化を用いること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（*A*）シグマ法による点数化、（*B*）1,2,3,4,5の配点からなる点数化、（*C*）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。

なお、リッカートが開発した尺度構成法では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．リッカート尺度に関する考察

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。

$$X_i = \frac{\sum_{s=1}^{m} X_{is}}{m} \qquad \text{または} \qquad X_i = \sum_{s=1}^{m} X_{is} \tag{2}$$

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。

先のセクションで示したようにリッカート（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、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。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場合である。

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でも、それは誤差を含んだものになる。そのような誤差について考察してみよう。

4．リッカート型項目データの誤差に関する考察

ここで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の誤差について考察する。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。

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

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W_iとu_iと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$W_j^*(j = 1,2,...,c)$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$X_j(j = 1,2,...,c)$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$W_j^*(j = 1,2,...,c)$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$\delta_j = X_j - W_j^*(j = 1,2,...,c)$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。

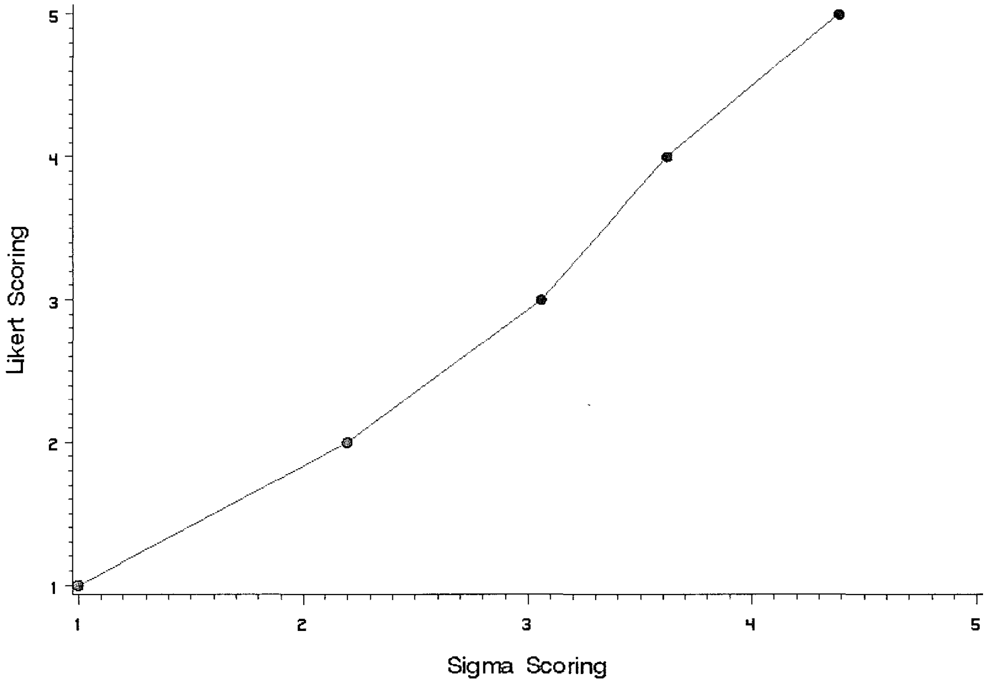

図２．簡便なリッカートの点数化を特徴づける図

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。

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$\overline{u}$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$\overline{u}$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$\overline{u}$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。

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

したがって誤差の平均$\overline{u}$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$\overline{u}$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$\overline{u} = 0$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$\overline{u}$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$\overline{u}$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。

5．リッカート型項目データを回帰に用いる場合の考察

ここではリッカート型項目データを回帰に用いる場合を考察する。第1の興味は通常最

小２乗推定量の不偏性にある。古典的線型回帰モデル[2)]において通常最小２乗推定量は最良線型不偏推定量となる。このため回帰においては通常最小２乗推定法が広く一般的に使われている。リッカート型項目データを説明変数側に用いても、通常最小２乗推定量の不偏性は保証されるのか。さらに一致性はあるのか。このようなことが主な興味である。

一般性を失うことなく具体的に議論を展開するため、真の回帰モデルは次のものであると仮定する。

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i \text{ with } \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \text{ for } i = 1,2,3,\ldots,n \qquad (8)$$

ただしY_iは従属変数（被説明変数）、X_{2i}とX_{3i}は説明変数、ε_iは撹乱項である。$\beta_1, \beta_2, \beta_3$はモデルの係数パラメータである。各説明変数と撹乱項とは独立しているなど、古典的線型回帰モデルの仮定[3)]を満たすものとする。

まず、説明変数にリッカート型項目データを用いる場合を考察する。説明変数X_{2i}が潜在変数であることから、そのためのリッカート型項目データX_{2i}^*を用いるものとし、次のように仮定する。

$$X_{2i}^* = X_{2i} + u_i \text{ with } u_i \sim N(0, \sigma_u^2) \text{ for } i = 1,2,\ldots,n \qquad (9)$$

ただしu_iはリッカート式の点数化に起因する観測の誤差を示す。異なる被験者の間で誤差項$u_i, u_j (i \neq j)$は互いに独立し、さらに誤差項u_iは他の説明変数X_{3i}や撹乱項ε_iとも独立しているとする。研究者が定めるリッカート型項目データのスケールによって係数パラメータの真の値が変化することに留意する必要がある。したがって研究者がリッカート型項目データX_{2i}^*のスケールを定めると、例えば点数配分を1,2,3,4,5に定めると、それに対応するスケールの潜在変数X_{2i}が特定され、それに応じて係数パラメータ$\beta_1, \beta_2, \beta_3$の真の値も定まる。しかし、リッカート型項目データ$X_{2i}^*$が誤差$u_i$を内包していることに変わりはない。

したがって実際の推定に用いられる回帰モデルは次のようになる。

$$Y_i = \beta_1 + \beta_2 (X_{2i}^* - u_i) + \beta_3 X_{3i} + \varepsilon_i$$

$$Y_i = \beta_1 + \beta_2 X_{2i}^* + \beta_3 X_{3i} + (-\beta_2 u_i + \varepsilon_i)$$

$$Y_i = \beta_1 + \beta_2 X_{2i}^* + \beta_3 X_{3i} + \varepsilon_i^* \qquad (10)$$

---

2) 説明変数が非確率変数か確率変数かによって古典的線型回帰モデルと新古典的線型回帰モデルといったように区別する場合もある。説明変数が確率変数の場合は説明変数のデータを所与とする条件での回帰となる。この点を除き、両モデルは実質的に同等であり、いずれの場合も通常最小２乗推定量は最良線型不偏推定量となる。したがって古典的線型回帰モデルとして統一している。

3) リッカート型項目データを説明変数に用いる場合は説明変数が確率変数であることを意味する。したがって説明変数のデータを所与とする条件での仮定となる。

ただし$\varepsilon_i^* = -\beta_2 u_i + \varepsilon_i$である。（１０）式が実際の推定に用いられる回帰モデルを示している。（9）式で示すようにX_{2i}^*とu_iとが相関していることから、（１０）式におけるX_{2i}^*とε_i^*とは相関する結果となる。このように説明変数と撹乱項とが相関している場合、係数パラメータの通常最小２乗推定量は不偏性を失い、一致性も失う。このような結論は回帰分析において「変数の誤差（errors in variables)」として知られている問題である。不偏性も一致性も無いような通常最小２乗推定に基づく回帰分析は、結局のところ信頼が置けない推定結果に基づく推論になる。さらにデータの分布が「正規性」を反映しておらず、間隔データとして取り扱うことができないようなリッカート型項目データの場合は、誤差の問題は極めて深刻になり、推論の信頼性はますます無くなる。

次は変数の誤差の問題に対処する推定法を考えてみよう。良く知られている推定法は操作変数法である。しかし、リッカート型項目データを用いる説明変数のための操作変数を見つけることは困難である。そのような操作変数があるならば、わざわざリッカート式アンケート調査をする必要が無いからである。そのような変数が無いからリッカート式アンケート調査を行なうのが普通である。

それではリッカート型項目データを回帰の説明変数側に用いることはあきらめなければならないのか。一つの対処法はダミー変数の使用にある。例えば反応カテゴリーが５個ある場合は４個のダミー変数を準備する。そして「強い賛成」の反応カテゴリーのためのダミー変数は「強い賛成」の回答に対して１、他の回答に対してはゼロの値をとるようにする。このようなことを他の３個のダミー変数についても行なう。反応カテゴリーの個数がc個の場合は$c-1$個のダミー変数を用いて対処する。このことはリッカート型項目データをデジタル化することを意味している。さらにリッカート型項目データを間隔データではなく、順位データとして取り扱っていることも意味している。このようにしてデジタル化したデータはリッカート式の点数化に起因する誤差が含まれていない。したがって変数の誤差の問題は回避できる。古典的仮定が崩れていないならば通常最小２乗推定量は最良線型不偏推定量となる。さらに通常最小２乗推定量は一致推定量となる。

次はリッカート型項目データを回帰の従属変数に用いる場合を考えてみよう。真の従属変数Y_iと、そのためのリッカート型項目データY_i^*とは次のようになっていると仮定する。

$$Y_i^* = Y_i + u_i \text{ with } u_i \sim N(0,\sigma_u^2) \text{ for } i = 1,2,...,n \quad (11)$$

ただしu_iはリッカート式の点数化に起因する観測の誤差を示す。ここでも異なる被験者の間で誤差項$u_i, u_j (i \neq j)$は互い独立しており、さらに誤差項u_iは各説明変数や撹乱項ε_iとも独立であるとする。この場合は次のようになる。

$$Y_i^* - u_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i + \varepsilon_i$$

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \beta X_{3i} + \varepsilon_i^* \quad (12)$$

ただし$\varepsilon_i^* = u_i + \varepsilon_i$である。（１２）式において各説明変数と撹乱項$\varepsilon_i^*$とは無相関であり、古典的仮定は崩れていない。このように従属変数のデータが観測誤差を含むものであっても、説明変数が誤差なしで観測されておれば古典的仮定が崩れることはなく、通常最小２乗推定量は最良線型不偏推定量となる。

データの分布が「正規性」を反映しておれば、リッカート型項目データを従属変数に使うことに全く問題が無いかというと、そうではなく、次のようなことを認識する必要がある。変数の誤差（errors in variables）として知られている問題では量的変数が誤差を含む形で観測されることを想定しており、そこで想定されている誤差のスケールとリッカート型項目データの誤差のスケールとは大きく異なり、後者の方がはるかに大きいと考えられる。誤差のスケールが大きいことは、回帰の分散が大きくなり、推論の精度が低いものになることを意味する。さらにデータの分布が「正規性」を反映していない場合は、リッカート型項目データを量的データとして使うのではなく順位データとして従属変数に使うことが考えられる。その場合は順位応答モデル（ordered-response model）といった選択肢がある。

６．おわりに

リッカート型項目データを検証しないまま量的データとして用いて通常最小２乗推定法で回帰を行ない、その結果を学術論文で報告している実証的研究に対し、まず疑問を投げかけた。そしてリッカート型項目データは誤差を本質的に含んでいることを示した。そのようなデータを回帰の説明変数側に用いると、通常最小２乗推定量は不偏性を失い、一致性も失うことになる。これは回帰において「変数の誤差（errors in variables）」として知られている問題でもある。したがってリッカート型項目データを回帰の説明変数側に用い、通常最小２乗推定法で推定している回帰分析は、信頼のできない推定結果に基づいて推論していることになる。さらにデータの分布が「正規性」を反映しておらず、間隔データとして取り扱うことができないようなリッカート型項目データの場合は、誤差の問題は極めて深刻になる。

変数の誤差の問題に対処する推定法としては操作変数法が良く知られているが、リッカート式アンケート調査を行なうような社会科学の分野では、操作変数を見つけること自体が困難である。そのような変数が無いからリッカート式アンケート調査を行なうのが普通である。

このような状況における一つの対処法はダミー変数の使用にある。つまり、リッカート式アンケート調査から得たデータをデジタル化する訳である。それはリッカート式アンケート調査の結果を順位データとして取り扱うことを意味する。それにより変数の誤差の問題は回避できる。古典的仮定が満たされるならば通常最小２乗推定量は最良線型不偏推定量となる。

（2003年12月18日受付、2003年12月24日受理）

参考文献

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.

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

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

Easthope, Gary (1974): *A History of Social Research Methods*, Longman.（阿久津 昌三, 他 4 名 訳（1982）: 社会調査方法史, 慶應通信）.

Garson, David G. (n.d.): Data Levels and Measurement (http://www2.chass.ncsu.edu/garson/pa765/datalevl.htm, 2003.6).

Goldberger, Arthur S. (1991): *A Course in Econometrics*, Harvard University Press.

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.

Greene, William M. (2000): *Econometric Analysis* (4th Edition), Prentice Hall.

Hayashi, Fumio (2000): *Econometrics*, Princeton University Press.

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.

Iman, Ronald L. and W. J. Conover (1979): The Use of the Rank Transform in Regression. *Techonometrics* 21 (4), 499-509.

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*, Archives of Psychology, New York.

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

Mohberg, N. R., M. Ghosh, and J. E. Grizzle (1978): Linear Models Analysis of Small Samples of Categorized Ordinal Response Data. *Journal of Statistical Computation and Simulation* 7, 237-251.

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.

Schmidt, Peter (1976): *Econometrics*, New York: Marcel Dekker.

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), Aldine, New York, 415-431.

安田 三郎, 原 順輔（1984）: 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-type surveys are widely used in the field of social science, and data of individual Likert-type items, called Likert-type item data in this paper, are used for regression analysis as well. Doubt is put on the reliability of such regression analysis. It is shown that Likert-type item data contain measurement errors, and the use of such data for explanatory variables in regression makes the ordinary least squares (OLS) estimator biased and inconsistent. This is the same as the problem of "errors in variables" in regression literature. The problem gets worse if Likert-type item data under consideration do not have qualification of interval data and such qualification is not checked by researchers.

To deal with the problem of "errors in variables," the popular estimation method is the instrument variable method. However, it is impossible to find instrument variables in the case of Likert-type surveys. A Likert-type survey is conducted since such variables are not available.

One method for dealing with this situation is the use of dummy variables. In other words, it is the digitalization of data obtained from a Likert-type survey. Such digital data do not contain quantitative measurement errors. Unless the classical assumptions are violated, the OLS estimator is the best unbiased linear estimator.