

# 共通&独自因子得点を説明変数に用いた重回帰分析

橋 本 翔

西南学院大学商学論集  
第71巻 第1・2合併号 抜刷  
2024（令和6）年 9 月 発行

# 共通&独自因子得点を説明変数に用いた重回帰分析

橋 本 翔

## はじめに

本研究で対象とする場面は、因子分析の共通因子得点と個々の項目がどれだけ目的変数に影響を及ぼしているか推定したい、という場面である。例えば、小売店で個々の食料品を客に評価してもらったデータをもとに因子分析を行ったところ「値段の安さ」「量の多さ」「味の美味しさ」を説明する共通因子として「お得感」という因子が得られたとしよう。そして、この場面において「商品の魅力」という項目とこれらの項目との関連に関心があったとしよう。このとき、「お得感」因子を統制したもとの、つまり同じお得感であれば、「値段の安さ」「量の多さ」「味の美味しさ」のそれぞれがどの程度「商品の魅力」に影響を及ぼすかを推定したいといった場面を本研究では取り扱う。前述のようなモデルから得られる知見は、商品開発の場面や消費者理解において役立つだろう。また、このように全体的な傾向を統制したもとの個々の項目の影響を知りたい、もしくは全体の傾向と個々の項目の影響度を比較したいというニーズは様々な場面において存在すると考えられる。例えば、目的変数と心理尺度との関係を重回帰分析により推定する場面は多いが、心理尺度を測定する複数の項目間には相関があり、未知の因果構造を持っている可能性があるため、重回帰分析の説明変数として用いても結果の解釈は困難である（吉田, 2021）。そのような状態では、全体の傾向を共通因子として統制した上での個々の項目が目的変数に与える影響は、興味の対象となり得る。このような目的のもと、本研究では重回帰分析の説明変数として共通因子得点と

独自因子得点の両方を用いる手法を提案する。

重回帰分析は予測や現象の記述、因果推論などに広く用いられている分析手法であり、統計的データ解析手法の中で最も多く応用されるきわめて有用な手法である（岩崎, 2021）。重回帰分析では複数の説明変数を用いて、単一の目的変数の値を予測することを目的としている。ここで、 $n$ 個体  $\times p$  変数の標準化された説明変数行列を  $\mathbf{X}$ 、 $n$ 個体  $\times 1$  変数の標準化された目的変数ベクトルを  $\mathbf{y}$  とすると、重回帰分析の最小二乗推定では、以下の目的関数の最小化を行う偏回帰係数ベクトル  $\beta$  の推定を行う。

$$l(\beta|\mathbf{X}, \mathbf{y}) = \|\mathbf{y} - \mathbf{X}\beta\|^2 \quad (1)$$

また、この目的関数の値を最小にする偏回帰係数ベクトルの推定値  $\hat{\beta}$  は次のように与えられる。

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2)$$

ここで、 $(\mathbf{X}'\mathbf{X})^{-1}$  は  $(\mathbf{X}'\mathbf{X})$  の逆行列を示す。重回帰分析は一般的には説明変数および目的変数が量的変数である場合のみならず、変数の変換やダミー変数を用いることも可能であり、その適応対象となる場面は非常に広い。例えば、説明変数が名義尺度変数であってもダミー変数を用いることで影響度を推定することができる。また、個々の説明変数に対する変数変換や、複数の説明変数による新たな説明変数の作成などは容易に行うことができる。また、各説明変数とそれに対応する重みで目的変数を予測するという明瞭なモデルは分析結果の解釈の上でも簡便である。これらの理由から、重回帰分析は様々な分野で用いられている。

また、因果推論の文脈では重回帰分析は交絡因子が存在する際に、その統制を行うことで因果効果を推定する手法として用いられる。交絡因子とは、説明変数と目的変数の両者に影響を与える変数であり、これが存在している場合には回帰分析で推定された回帰係数は本来の因果効果を表せていない場合がある。重回帰分析で推定される偏回帰係数は「ある説明変数が1単位変動した時に、“他の変数が同一である条件での、”目的変数の推定値（予測値）の変化量」を示す。そのため、興味の対象である変数と同時に交絡変数を説明変数に含めることにより、交絡変数の影響を統制した

うえて、興味の対象である変数が目的変数に与える効果を推定することができる。例えば、若年世代の年収の決定要因を調べる上で、教育年数の年収への効果を推定する際には、年齢が交絡となりうる。教育年数とともに年齢を説明変数として用いることにより、年齢の影響を取り除いたもとの教育年数から年収への効果を推定することができる。

本研究は、因子分析 (Factor Analysis: FA) で得られる潜在変数である共通因子と因子の推定に用いた顕在変数のそれぞれが目的変数に与える影響を推定することを目的にしている。因子分析では、互いに相関を持つ複数の顕在変数の共変同を説明するために、それらの変数の背後に共通の要因となる潜在変数が存在することを仮定し、潜在変数から各顕在変数への影響の大きさを求める。因子分析の定式化はしばしば相関行列へのフィッティングに基づいてなされるが、そうではなく因子得点も母数として推定するモデルとしては行列分解型の因子分析 (de Leeuw, 2004; Unkel and Trendafilov, 2010; Adachi, 2012) が存在している。ここで  $n$  個体  $\times p$  変数の標準化された顕在変数行列を  $\mathbf{X}$  とすると、因子数  $m < p$  の行列分解型因子分析では次の目的関数を最小にするパラメータ  $\mathbf{A}, \mathbf{F}, \mathbf{\Psi}, \mathbf{U}$  を求めることを目指す。

$$l(\mathbf{A}, \mathbf{F}, \mathbf{\Psi}, \mathbf{U} | \mathbf{X}) = \|\mathbf{X} - (\mathbf{F}\mathbf{A}' + \mathbf{U}\mathbf{\Psi})\|^2 \quad (3)$$

ここで、 $\mathbf{F}$  は  $n$  個体  $\times m$  因子の共通因子得点行列、 $\mathbf{A}$  は  $p$  変数  $\times m$  因子の共通因子負荷行列、 $\mathbf{U}$  は  $n$  個体  $\times p$  変数の独自因子得点行列、 $\mathbf{\Psi}$  は  $p$  変数  $\times p$  変数の対角行列である独自因子負荷行列である。ただし、上記のパラメータには以下の制約がある。

$$\mathbf{F}'\mathbf{1} = \mathbf{0}, \mathbf{F}'\mathbf{F} = n\mathbf{I}_m, \mathbf{U}'\mathbf{1} = \mathbf{0}, \mathbf{U}'\mathbf{U} = n\mathbf{I}_p, \mathbf{F}'\mathbf{U} = \mathbf{0}_{m \times p} \quad (4)$$

これらの制約はそれぞれ、因子得点は平均が 0 であり分散が 1 であること、共通因子得点間の無相関性、独自因子得点間の無相関性、共通因子得点と独自因子得点の無相関性の仮定を示している。

因子分析が使用される場面は多岐に渡っている。例えば、マーケティングにおけるプリファレンスマッピングの作成では、製品の主観評価データをもとに因子分析を行い、各製品を 2 つの因子得点を用いてマッピングす

るという手法がとられている。また、アンケート調査で心理変数や行動変数を用いて性格や嗜好などの個人特性を調査する場面においても、個々の項目に対して因子分析を用いて因子得点を推定し、因子得点を用いて個人を分類したり、因子得点と興味の対象となる変数との関係を推測したりする。

しかしながら、共通の傾向を統制するために因子得点と個々の項目を同時に重回帰分析の説明変数として用いることは困難である。これは、個々の項目の得点から因子得点を算出し、それと各項目の得点を同時に重回帰分析の説明変数に用いるという素朴なアプローチには、説明変数間に完全な多重共線性が存在するという問題があるためである。これは一般的に因子得点が顕在変数行列の線形変換によって定められる点に起因する。因子得点には因子負荷行列が定まったもとでも、推定値が一意に定まらないという性質が存在している。そのためデータおよび因子負荷行列から共通因子得点を得るにはいくつかの手法があり、代表的な手法としては、回帰法やBartlett法 (Bartlett, 1937) などがあげられる。これらのいずれの手法も得られる因子得点はデータ行列の線形変換によって定められる。そのため下位項目と因子得点を同時に説明変数に用いた重回帰分析を行った際には、説明変数間に線形関係があるため一意な解を求めることができない。無論、これは心理尺度を用いた際の尺度得点の算出の場合も同様である。尺度得点の算出の際には平均や合計が用いられることが多いが、これは特定の因子負荷行列を与え、回帰法で因子得点を求めることと同義である。これらの理由によって、因子得点とそれを求めた顕在変数を同時に説明変数として用いた際には、多重共線性の問題が生じる。

重回帰分析において多重共線性への対処としては、一般に次の3つの対処法がとられる。1つ目は説明変数の削減である。削減する変数はドメイン知識を利用して理論的に導く、個々の説明変数のVIF (Variance inflation factor) を算出し値の高いものを削減する、赤池情報量基準 (Akaike Information Criterion: AIC; Akaike, 1973) などのモデル選択基準をもとにステップワイズ法などの変数選択法を用いるなどの手法が存在する。2つ目

は説明変数の合成である。これは一般に多重共線性が存在する場合には説明変数間に高い相関を持つ項目が多いため、それらの変数を合成し（例えば合計し）、変数を置き換えるというものである。多重共線性の原因が3変数以上の変数群の場合には、それらに対して主成分分析を行い、第1主成分得点を新たな説明変数として導入し、主成分に関連の高い変数を取り除くということも行われる。3つ目は正則化項（罰則項）の導入である。重回帰分析では式(1)を最小化するパラメータ $\beta$ を求めるが、このアプローチでは

$$l^*(\beta|\mathbf{X}, \mathbf{y}) = \|\mathbf{y} - \mathbf{X}\beta\|^2 + f(\beta) \quad (4)$$

を最小にする $\beta$ を求める。これは、完全な多重共線性の数理的な問題は式(2)において $(\mathbf{X}'\mathbf{X})^{-1}$ の値が計算できないことに起因するため、この計算を行わなくてよい推定値を求めるというアプローチである。この時の正則化項 $f(\beta)$ の定め方によってさまざまな手法があるが、代表的なものは $f(\beta) = \lambda\|\beta\|^2$ とするRidge回帰（Hoerl & Kennard, 1970）と、 $f(\beta) = \lambda|\beta|$ とするLasso回帰（Tibshirani, 1996）である。なお、 $\lambda > 0$ は正則化係数であり、分析者が定める。

上記で挙げた3つの手法は、因子得点と個々の変数の効果を同時に推定するという目的に用いるのは困難である。変数削減の方法では、全ての説明変数が含まれたモデルにおけるVIFは計算することができず、またある説明変数が削減されたことは当該の変数が目的変数に影響を与えないことが示されたわけではない。また、変数統合の方法においても、変数を統合して得られたのが因子得点であるため、それをさらに統合するのは不合理である。また正則化項を用いる手法では、回帰係数を縮小推定するバイアスが生じる。

そこで本研究では所与の因子負荷行列のもとで、説明変数を各変数に影響を与える共通因子得点と個々の変数に影響を与える独自因子得点に分解し、それらを説明変数に用いて単一の目的変数を回帰する重回帰分析モデルを提案する。顕在変数である説明変数は目的変数に直接影響を与えているわけではなく、顕在変数の変動の背後に共通因子と独自因子が存在し目

の変数はそれらの因子から影響を受けているとみなす。このとき、式(4)で示されているように共通因子得点と独自因子得点は無相関であるため、説明変数間の線形性に起因する多重共線性の問題は生じない。

本論文の以下は次のように構成されている。次の章では、提案手法を用いたパラメータの推定について述べる。次に、数値シミュレーションによって提案手法と上述した正則化法とをパラメータの再現性の観点から比較する。その後、実データに対して適用し、提案手法の有用性について検討する。最後に、それらの結果について考察を行う。

### 提案手法

本研究においては共通因子の数は1つと定め、以下の目的関数の最適化について考える。また、議論の簡便のために $\mathbf{X}$ ,  $\mathbf{y}$ は標準化されているものとする。

$$l(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) = \|\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}\|^2 \quad (5)$$

$$\mathbf{Z} = (\mathbf{f}, \mathbf{U}) \quad (6)$$

なお、 $\mathbf{f}$ は共通因子得点行ベクトル、 $\mathbf{Z}$ は共通因子得点行列 $\mathbf{f}$ と独自因子得点行列 $\mathbf{U}$ を列方向に結合した $n \times (p+1)$ の因子得点行列である。また、偏回帰係数ベクトル $\boldsymbol{\beta}$ は $(p+1)$ 次の行ベクトルである。なお、因子負荷行列は因子数が1であるため行ベクトルとなるが、ここでは記述の簡便のために $p \times 1$ の行列として扱う。ここで顕在変数は標準化されている前提にしたが、因子分析は尺度不変性という性質があるため結果には大きく影響しない。

まず、因子負荷行列 $\mathbf{A}$ および独自因子負荷行列 $\boldsymbol{\Psi}$ が得られた状態において $\mathbf{Z}$ は以下のように表現することが可能である (Unkel and Trendafilov, 2010; Adachi, 2013; Uno et. al., 2019)。

$$\mathbf{Z} = \sqrt{n}(\mathbf{K}_1 \mathbf{L}'_1 + \mathbf{k}_2 \mathbf{l}'_2) \quad (7)$$

ここで、 $\mathbf{K}_1, \mathbf{L}_1, \mathbf{K}_2, \mathbf{L}_2$ は以下の特異値分解

$$\mathbf{X}\mathbf{A} = (\mathbf{K}_1, \mathbf{k}_2) \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} (\mathbf{L}_1, \mathbf{L}_2)' \quad (8)$$

$$\mathbf{A} = (\mathbf{A} \ \boldsymbol{\Psi}) \quad (9)$$

の左特異行列ならびに右特異行列から得られる。ここで、 $\mathbf{K}_1, \mathbf{L}'_1$ はそれぞれ左特異ベクトル行列および右特異ベクトル行列の第1列から第 $p$ 列までの行列、 $\mathbf{K}_1, \mathbf{L}'_1$ はそれぞれ左特異ベクトル行列および右特異ベクトル行列の第 $(p+1)$ 列、 $\mathbf{D}_1$ は $\mathbf{X}\mathbf{A}$ の特異値が左上から右下に降順で並んだ $p \times p$ の対角行列であり、 $\mathbf{A}$ は共通因子負荷行列 $\mathbf{\Lambda}$ と独自因子負荷行列 $\mathbf{\Psi}$ を列方向に並べたものである。また上記の特異値分解は $\mathbf{K}_1, \mathbf{L}'_1$ については単一の値が得られるが、 $\mathbf{k}_2, \mathbf{l}_2$ は以下の制約

$$\mathbf{K}'_1 \mathbf{k}_2 = \mathbf{0} \tag{10}$$

$$\mathbf{k}'_2 \mathbf{k}_2 = 1 \tag{11}$$

$$\mathbf{L}'_1 \mathbf{l}_2 = \mathbf{0} \tag{12}$$

$$\mathbf{l}'_2 \mathbf{l}_2 = 1 \tag{13}$$

$$\mathbf{Z}'\mathbf{Z} = n\mathbf{I}_{p+1} \tag{14}$$

を満たす任意の値でよい。ただし $\mathbf{l}_2$ に関しては、式(14)および式(7)(10)より、

$$\sqrt{n}(\mathbf{K}_1 \mathbf{L}'_1 + \mathbf{k}_2 \mathbf{l}'_2)' \sqrt{n}(\mathbf{K}_1 \mathbf{L}'_1 + \mathbf{k}_2 \mathbf{l}'_2) = n\mathbf{I}_{p+1} \tag{15}$$

$$\mathbf{L}_1 \mathbf{L}'_1 + \mathbf{l}_2 \mathbf{l}'_2 = \mathbf{I}_{p+1} \tag{16}$$

$$\mathbf{l}_2 \mathbf{l}'_2 = \mathbf{I}_{p+1} - \mathbf{L}_1 \mathbf{L}'_1 \tag{17}$$

であることが定まっており、 $\mathbf{I}_{p+1} - \mathbf{L}_1 \mathbf{L}'_1$ の階数は1であり、またその固有値は1であるため、 $\mathbf{l}_2$ は $\mathbf{I}_{p+1} - \mathbf{L}_1 \mathbf{L}'_1$ の固有値分解の第一固有ベクトルとして得られる。ただし、符号の不定性は有する。なお、 $\mathbf{0}$ は0ベクトル、 $\mathbf{I}_{p+1}$ は $(p+1)$ 次の単位行列を示す。

さて、式(7)を用いると式(5)は次のように書くことができる。

$$l(\boldsymbol{\beta}, \mathbf{k}_2 | \mathbf{X}, \mathbf{y}) = \|\mathbf{y} - \sqrt{n}(\mathbf{K}_1 \mathbf{L}'_1 + \mathbf{k}_2 \mathbf{l}'_2) \boldsymbol{\beta}\|^2 \tag{18}$$

以降では、式(18)を最適にする $\boldsymbol{\beta}, \mathbf{k}_2$ について考える。

### $\boldsymbol{\beta}$ の最適化

まず本節では、 $\mathbf{k}_2$ を固定したもつで $\boldsymbol{\beta}$ を最適化することを考える。これは説明変数が $\sqrt{n}(\mathbf{K}_1 \mathbf{L}'_1 + \mathbf{k}_2 \mathbf{l}'_2)$ 、目的変数が $\mathbf{y}$ の重回帰分析の偏回帰係数を求めることを意味するので、式(2)(7)(14)より



$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \sqrt{n}^{-1}(\mathbf{K}_1\mathbf{L}'_1 + \mathbf{k}_2\mathbf{l}'_2)\mathbf{y} \quad (19)$$

で得られる。これを展開すると、

$$\hat{\boldsymbol{\beta}} = \sqrt{n}^{-1}\mathbf{L}_1\mathbf{K}'_1\mathbf{y} + \sqrt{n}^{-1}\mathbf{l}_2\mathbf{k}'_2\mathbf{y} \quad (20)$$

となり、データから定まる部分（第1項）と制約を用いて定める部分（第2項）に分解できる。

ここで式(10)より $\mathbf{J}_{\mathbf{K}_1} = (\mathbf{I}_n - \mathbf{K}_1\mathbf{K}'_1)$ とおくと、 $\mathbf{k}_2 = \mathbf{J}_{\mathbf{K}_1}\mathbf{k}_2$ が成り立てば制約(10)が満たされることがわかる。 $\mathbf{J}_{\mathbf{K}_1}\mathbf{y}$ の標準偏差は $\sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}}$ であるため、

$$\mathbf{k}'_2\mathbf{y} = \mathbf{k}'_2\mathbf{J}_{\mathbf{K}_1}\mathbf{y} = \mathbf{k}'_2\mathbf{J}_{\mathbf{K}_1}\mathbf{y} \left( \sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}} \right)^{-1} \left( \sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}} \right) = \sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}} \times r_{\mathbf{k}_2\mathbf{y}} \quad (21)$$

となる。ここで、 $r_{\mathbf{k}_2\mathbf{y}} = \mathbf{k}'_2\mathbf{J}_{\mathbf{K}_1}\mathbf{y} \left( \sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}} \right)^{-1}$ は $\mathbf{k}_2$ と「 $\mathbf{y}$ の $\mathbf{k}_1$ で説明できない部分」との相関係数である。以上より、 $\boldsymbol{\beta}$ の推定値は $\sqrt{n}^{-1}\mathbf{L}_1\mathbf{K}'_1\mathbf{y}$ を基準に $\sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}/n} \times \mathbf{l}_2 \times r_{\mathbf{k}_2\mathbf{y}}$ だけ変動した値だと捉えることができる。

## $\mathbf{k}_2$ の最適化

本節では $\hat{\boldsymbol{\beta}}$ を固定したもとの、最適な $\mathbf{k}_2$ を求めることを考える。式(18)は式(10)～(14)の制約から

$$\|\mathbf{y} - \sqrt{n}(\mathbf{K}_1\mathbf{L}'_1 + \mathbf{k}_2\mathbf{l}'_2)\hat{\boldsymbol{\beta}}\|^2 = -2\sqrt{n}\hat{\boldsymbol{\beta}}'\mathbf{l}_2\mathbf{k}'_2\mathbf{J}_{\mathbf{K}_1}\mathbf{y} + const \quad (22)$$

と整理できる。なお、 $const$ は最適化に関係しない定数である。ここで、 $\hat{\boldsymbol{\beta}}'\mathbf{l}_2$ は正のスカラーなので式(20)の最小化は $\mathbf{k}'_2\mathbf{J}_{\mathbf{K}_1}\mathbf{y}$ の最大化によってなされることが分かる。これを制約(10)(11)のもとで達成する $\mathbf{k}'_2$ はTemBerge(1993)より、

$$\mathbf{k}_2 = \frac{\mathbf{J}_{\mathbf{K}_1}\mathbf{y}}{\sqrt{\mathbf{y}'\mathbf{J}_{\mathbf{K}_1}\mathbf{y}}} \quad (23)$$

として得られる。

以上の手続きにより式(20)を最小にする $\mathbf{k}_2$ は得られたが、最適な $\mathbf{k}_2$ を定めるといことは、分析者が説明変数を都合よく操作していることと同義である。次の節では、この操作において分析者がどのような仮定をおいていることになるのかを明示化するために、 $\mathbf{k}_2$ が示す情報を相関係数 $r_{\mathbf{k}_2\mathbf{y}}$ を

通して考える。

### 相関係数 $r_{k_2y}$ の解釈と $\beta$ の範囲

相関係数 $r_{k_2y}$ は $k_2$ と「 $y$ の $X$ で説明できない部分」との相関を表したものであるが、これを定めることは分析者が何を仮定していることを意味するだろうか。これを明らかにするため、 $k_2$ がどのような情報であるかについて考える。まず式(8)より $XA = K_1 D_1 L_1$ であるので式(10)から $X'k_2 = 0$ つまり説明変数行列と $k_2$ は無相関である。また、式(7)および式(10)(11)から $Z'\sqrt{n}k_2/n = l_2$ となるため、各因子との相関係数は $l_2$ であり相関がある。つまり、 $k_2$ は説明変数とは独立であるが、共通因子・独自因子とは関連しているような、各個体に与えられる情報であると言える。よって相関係数 $r_{k_2y}$ は「説明変数とは無相関であるが、共通因子・独自因子とは関連している情報と目的変数との相関関係の強さ」を意味する。相関係数は通常 $-1 \sim +1$ の範囲の値をとるが、この値の絶対値を1と定めることは、「説明変数とは無相関である共通因子・独自因子に影響を与えている外部要素が、目的変数を最大限説明している」状況を仮定していることになる。逆に0と定めると「説明変数とは無相関の共通因子・独自因子に影響する外部要素は、目的変数と無相関である」状況を仮定していることになる。このどちらの仮定も極端な例ではあるが、モデルが正しいのであればこれらの2つの状況の間に真のパラメータが存在するとみなすことができる。

分析者の立場によって適切な $r_{k_2y}$ の設定は異なる。前述のように目的関数の最小化、つまり目的変数を最大限説明するように設定するという観点では $r_{k_2y} = 1$ あるいは $r_{k_2y} = -1$ と定めるのが最適になる。一方、説明変数以外の情報は利用すべきではないという立場では、 $r_{k_2y} = 0$ と定めるのが適当だろう。後者の場合は、偏相関係数の推定値は一意に定まる一方、前者の場合には $\beta$ の点推定値は一意に定めることが困難である。偏相関係数 $\beta$ は定数である $\sqrt{n}^{-1}L_1K_1'y$ を基準に $\sqrt{y'K_1y/n} \times l_2 \times r_{k_2y}$ だけ変化した値である。なお、 $l_2$ には符号の不定性が存在するため、

$$\hat{\beta} = \sqrt{n}^{-1} \mathbf{L}_1 \mathbf{K}'_1 \mathbf{y} + \sqrt{\frac{\mathbf{y}' \mathbf{J}_{\mathbf{K}_1} \mathbf{y}}{n}} \times l_2 \times r_{k_2 y} \quad (24)$$

$$\hat{\beta} = \sqrt{n}^{-1} \mathbf{L}_1 \mathbf{K}'_1 \mathbf{y} - \sqrt{\frac{\mathbf{y}' \mathbf{J}_{\mathbf{K}_1} \mathbf{y}}{n}} \times l_2 \times r_{k_2 y} \quad (25)$$

のどちらもがその解となる。これら2つの解はデータとの適合度という点ではどちらも等しいため、 $r_{k_2 y}$ の値を固定しても偏回帰係数 $\beta$ の点推定値をそのまま利用することは困難である。

そのため本論文では $\sqrt{n}^{-1} \mathbf{L}_1 \mathbf{K}'_1 \mathbf{y} \pm \sqrt{\frac{\mathbf{y}' \mathbf{J}_{\mathbf{K}_1} \mathbf{y}}{n}} \times l_2 \times r_{k_2 y}$ を $\beta$ の取りうる値の範囲として解釈することを提案する。つまり、分析者が相関係数 $r_{k_2 y}$ の最大値を $r$ と仮定した場合、因子モデルが正しいとすると偏回帰係数 $\beta$ の値は $\sqrt{n}^{-1} \mathbf{L}_1 \mathbf{K}'_1 \mathbf{y} \pm \sqrt{\frac{\mathbf{y}' \mathbf{J}_{\mathbf{K}_1} \mathbf{y}}{n}} \times l_2 \times r_{k_2 y}$ の $r_{k_2 y}$ を $-1$ から $+1$ まで変動させた範囲の中に存在するとみなせる。この範囲にもとづいて個々の因子が目的変数に与える影響を観測することができる。例えばある独自因子が取りうる値の範囲の中に0を含まないのであれば、その独自因子は真の相関係数 $r_{k_2 y}$ がどのような値であったとしても、目的変数の予測に利用できるとみなすことができる。

また、この範囲の中の値について、何らかの仮説がある場合にはそれを利用した解を求めることもできる。例えば、一つまたは複数の独自因子の偏回帰係数が0であると仮定できる、つまりいくつかの独立変数からの影響は0であると仮定できるのであれば、それもしくはそれらの係数が0に最も近くなる $r_{k_2 y}$ の値を定めれば良い。

このように事前情報を用いて一意な解を用いた際や $r_{k_2 y} = 0$ と定めた際には一意な偏回帰係数の推定値が得られるが、この時には偏回帰係数の区間推定を行うことができる。ある偏回帰係数 $\beta_j$ の推定値 $\hat{\beta}_j$ の従う分布は説明変数となる因子得点行列 $\mathbf{Z}$ が $(\mathbf{Z}'\mathbf{Z})^{-1} = n^{-1}\mathbf{I}$ であるため、残差に正規性を仮定し分散を $\sigma^2$ と表すと、 $\hat{\beta}_j \sim N(\beta_j, \sigma^2 n^{-1})$ と表せる。よって、

$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 n^{-1}}} \sim N(0, 1)$  となり、この  $\sigma^2$  を  $\hat{\sigma}^2 = \|\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}\|^2 / (n - (p + 1) - 1)$  で置き換えた  $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 n^{-1}}}$  は自由度  $n - p - 2$  の  $t$  分布に従う。これを利用して95%信頼区間を求めると、

$$\hat{\beta}_j + \sqrt{\hat{\sigma}^2 n^{-1}} t_{n-p-2}(0.025) < \beta_j < \hat{\beta}_j + \sqrt{\hat{\sigma}^2 n^{-1}} t_{n-p-2}(0.975) \quad (26)$$

と表すことができる。なおここで、 $t_d(a)$  は自由度が  $d$  の  $t$  分布における quantile function (累積確率が  $a$  となる点を返す関数) である。

次章では、相関係数  $r_{k_2y} = 0$  と定めたときの推定値の性質、推定値の範囲の性質と「ある独自因子の偏回帰係数は0である」という仮説を利用した推定値の性質のそれぞれについて数値シミュレーションを用いて示す。

## 数値シミュレーション

### 数値シミュレーション1：真の偏回帰係数の値に知見がない場合

#### 方法

提案手法の分析結果の妥当性を検証するために、人工的にデータを生成し、パラメータの推定精度について検証を行った。推定精度に関わると考え操作した要因は、サンプルサイズ・説明変数の数( $p+1$ )・因子分析モデルにおける誤差の大きさ・回帰分析モデルにおける誤差の大きさの4つである。それぞれ以下のように条件を定めた。

1. サンプルサイズ：4条件 ( $n = 50, 100, 150, 200$ )
2. 説明変数の数：3条件 ( $p = 5, 10, 15$ )
3. 因子分析モデルにおける誤差：3条件 ( $\alpha_x = 0, 0.1, 0.3, 0.5$ )
4. 回帰モデルにおける誤差：3条件 ( $\alpha_y = 0, 0.1, 0.3, 0.5$ )

条件の組み合わせ ( $4 \times 3 \times 3 \times 3$ ) の各条件のそれぞれについて、以下の手順でパラメータの再現性を調査した。

#### 1. 真値の設定

まず、推定対象であるパラメータの真値を設定した。真の偏回帰係数  $\boldsymbol{\beta}^*$  の値は  $-1 \sim 1$  の範囲で乱数を用いて定め、偏回帰係数の2乗和が1になるように個々の値を2乗和の平方根で除した。また、真の因子負荷行

列 $\mathbf{A}^*$ は全変数の共通性が0.8になるように、以下の通りに定めた。

$$\mathbf{A}^* = \begin{pmatrix} \sqrt{0.8} & \sqrt{(1-0.8)} & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ \sqrt{0.8} & 0 & 0 & \sqrt{(1-0.8)} \end{pmatrix}$$

## 2. データの生成

分析対象となる説明変数群 $\mathbf{X}$ と目的変数 $\mathbf{y}$ を次のように作成した。

$$\begin{aligned} \mathbf{X} &= (1 - \alpha_x)^2 (\mathbf{F}, \mathbf{U}) \mathbf{A}^* + \alpha_x^2 \mathbf{E}_x \\ \mathbf{y} &= (1 - \alpha_y)^2 (\mathbf{F}, \mathbf{U}) \boldsymbol{\beta}^* + \alpha_y^2 \mathbf{e}_y \end{aligned}$$

なお、 $\alpha_x$ および $\alpha_y$ はデータ生成における誤差の大きさを定める定数であり、 $\mathbf{E}_x$ および $\mathbf{e}_y$ はそれぞれ説明変数と目的変数の誤差項である。

データ生成の手続きとしては、まず潜在変数である因子得点 $\mathbf{F}$ および $\mathbf{U}$ 、また誤差行列 $\mathbf{E}_x$ および $\mathbf{e}_y$ を(4)式の制約のもとで作成した。具体的には、行列 $\mathbf{F}$ ,  $\mathbf{U}$ ,  $\mathbf{E}_x$ およびベクトル $\mathbf{e}_y$ は多変量正規分布 $MN(\mathbf{0}, \mathbf{I}_{p+1+p+1})$ から平均が $\mathbf{0}$ でそれぞれが互いに相関を持たないように、統計分析環境R (R Core Team, 2023)のパッケージMASS (Venables & Ripley, 2002)の関数`mvrnorm`を用いて平均および共分散行列が真値と一致するようにデータを生成した。

## 3. 因子負荷行列の推定

まず、データ行列 $\mathbf{X}$ に対して因子分析を行い、因子負荷行列を求めた。具体的には、Rの`psych`パッケージ (Revelle, 2023)の関数`fa`を用い、因子数を1に設定し共通因子の因子負荷行列の推定値 $\hat{\boldsymbol{\lambda}}$ を求めた。その後、得られた $\hat{\boldsymbol{\lambda}}$ を用いて、各変数について独自性 $1 - \hat{\lambda}_j^2$  ( $j = 1, \dots, p$ )を求め、それらの正の平方根を対角成分に持つ対角行列を独自因子の負荷行列の推定値 $\hat{\boldsymbol{\Psi}}$ とした。

## 4. 偏回帰係数の推定

式(20)により相関係数 $r_{k_2y} = 0$ と定めた点推定値を求め、式(24)(25)により推定値の範囲を求めた。

### 5. パラメータの再現性の確認

パラメータの再現性の指標として、偏回帰係数について真値との平均二乗誤差 (Mean Squared Error: MSE) およびコサイン類似度 (Cosine Similarity: COS) を求めた。MSEは真値との差の大きさを示しており、この値が小さいほど真値の再現性がよいと判断できる。また、コサイン類似度は $-1 \sim 1$ の値を取り、1に近いほど真値を再現できていることになる。両者の違いとしては、MSEはスケールの違いに大きく影響されるが、COSは影響されない点がある。なお、MSEおよびコサイン類似度は以下のように計算される。

$$MSE = \frac{1}{p+1} \|\beta^* - \hat{\beta}\|^2$$

$$COS = \frac{\beta^{*T} \hat{\beta}}{\|\beta^*\| * \|\hat{\beta}\|}$$

以上の1～5の手続きを各条件について100回ずつ行い、MSEの平均値を各条件のパラメータの再現性とした。推定精度の比較対象としては、共通因子得点を回帰法および Bartlett 法のそれぞれで推定し、共通因子得点と説明変数行列を同時に用いた回帰モデルを Ridge 回帰と Lasso 回帰のそれぞれで推定した。これらの比較対象としているモデルは、厳密には推定対象はデータ生成モデルと異なるが、共通因子の統制を行ったうえで各説明変数の影響度を推定することができるという点で、提案手法と目的は一致している。

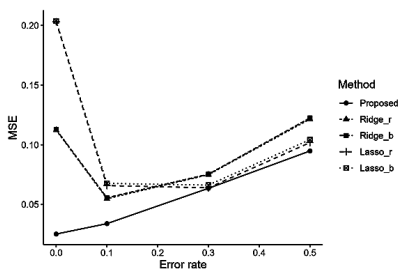


図1 手法ごとのエラーレート別 MSE

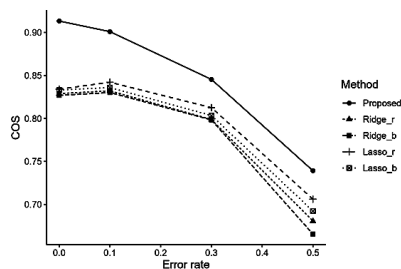


図2 手法ごとのエラーレート別 COS

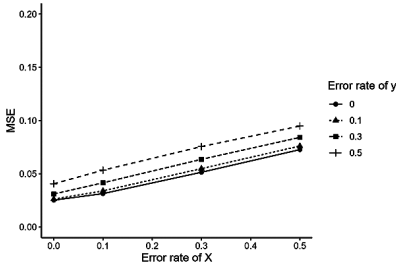


図3 エラーレート別 MSE

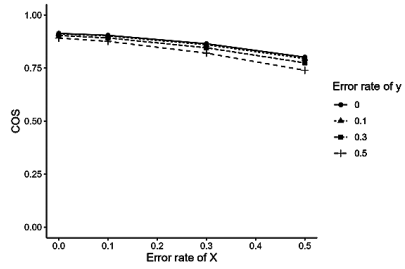


図4 エラーレート別 COS

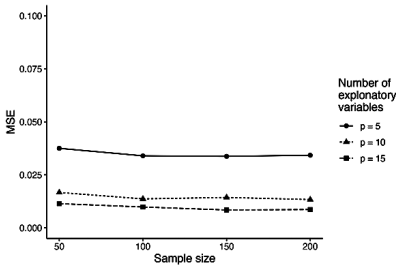


図5 サンプルサイズ×説明変数の数別 MSE

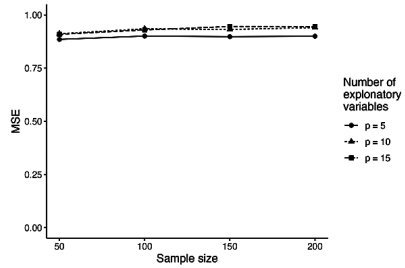


図6 サンプルサイズ×説明変数の数別 COS

### 結果

まず、他手法との再現性の比較を行い、提案手法の多手法に対する優位性を示す。図1は提案手法が他の方法に対して優位性を持つかを確認するために、サンプルサイズと説明変数の数をそれぞれ100と5に固定した状態でエラーレートを $\alpha_X = \alpha_Y$ の条件で変動させたときの、MSEの平均値を比較したものである。エラーレートがいずれの条件下においても、Ridge回帰とlasso回帰に対して提案手法のMSEの値は低いことが示されている。また、特にエラーレートが低い条件においては、提案手法のMSEの値が低い。また、図2はCOS類似度の値を示したものであるが、こちらも全ての手法に対して提案手法のCOS類似度の値が高い。これらの結果から、偏回帰係数の推定における提案手法の優位性が示された。また、Ridge回帰とLasso回帰のMSEの値は因子得点を回帰法で求めるかBartlett法で求めるか

にはほぼ依存しないことも明らかになった。一方、COSの値は両手法とも回帰法を用いた推定結果の方が推定精度が良いことがわかる。なお、本稿では省略するが、サンプルサイズ・変数の数が異なる状況でもこの傾向は同一であった。

次に、提案手法における回帰モデルにおけるエラー( $\mathbf{e}_y$ )の大きさと因子分析モデルにおけるエラー( $\mathbf{E}_X$ )の大きさが真値の再現性に影響を及ぼすのかを検証する。図3はそれぞれのエラーレートの値を変化させたときのMSEの値を示している。なお、各条件についてサンプルサイズは $n=100$ 、説明変数の数は $p=5$ で同一である。これを見ると、まず、XとYの誤差は両方ともに増大するとMSEの値が高くなっている。またそれぞれのエラーレートの影響を見ると、説明変数X側のエラーレートの大きさよりも目的変数Y側のエラーレートの方が変動に伴うMSEの変化量が大きいことがわかる。また、各グラフが並行関係であることから、エラーレートの間に交互作用はないことがわかる。一方で、図4で示されているCOS類似度の値は目的変数側の説明変数のエラーレートの大きさによる影響は目的変数側のエラーレートの大きさが大きいほど大きくなっている。これらのことから、因子分析モデルにおける誤差の大きさの影響と回帰分析における誤差の大きさのそれぞれが再現性に与える影響は、それぞれ誤差が増大すると再現性が低下すること、目的変数側の誤差の大きさがより大きく影響すること、またそれらには絶対値の大きさの面では交互作用がないが、ベクトルの類似性としては交互作用があることが分かる。

第3に、サンプルサイズと説明変数の数が推定に与える影響について確認する。図3はサンプルサイズと説明変数が異なる条件のMSEの値を比較したものである。なお、エラーレートはX、Yともに $a_X = a_Y = 0.1$ で同一である。図3からこのエラー条件の元ではサンプルサイズについては再現性に対する影響は大きくなく、100程度あれば十分であることが分かる。また、説明変数の数についても $p=10$ の条件と $p=15$ の条件で再現性に差はないことから、説明変数は10程度あれば十分に推定精度を保てることを見て取れる。



最後に推定された範囲の有効性について示す。表1は式(26)で示される範囲が0を含まなかった場合に、真の偏回帰係数の符号と計算された範囲で最も0に近い値の符号とが合致していた割合を示している。なお、エラーレートはX、Yともに $a_x = a_y$ の条件で0から0.5まで変動させている。この表から、エラーレートが0の場合にはすべてのサンプルサイズ、変数の数で1になっており、完全に再現出来ていることがうかがえる。また、 $n \geq 100$ の条件では、 $p \leq 10$ で完全に再現出来ており、また、 $n = 200$ の条件では $p = 15$ の条件でも符号の推定における範囲の情報の有用性が示されている。これらのことからサンプルサイズが十分に大きければ、偏回帰係数の符号の推定において、計算された範囲の情報は有用であることが示された。

表1 推定された範囲内に真値が含まれている割合

		error rate			
		0	0.1	0.3	0.5
n = 50	p = 5	1.00	0.98	0.98	0.96
	p = 10	1.00	0.99	0.97	0.87
	p = 15	1.00	0.99	0.96	0.85
n = 100	p = 5	1.00	1.00	1.00	1.00
	p = 10	1.00	1.00	1.00	1.00
	p = 15	1.00	1.00	0.99	0.93
n = 150	p = 5	1.00	1.00	1.00	1.00
	p = 10	1.00	1.00	1.00	1.00
	p = 15	1.00	1.00	1.00	0.97
n = 200	p = 5	1.00	1.00	1.00	1.00
	p = 10	1.00	1.00	1.00	1.00
	p = 15	1.00	1.00	1.00	1.00

### シミュレーション 2：真の偏回帰係数の0の位置が明らかな場合

次に、0の位置についての知見がある場合の、推定値の精度に関するシミュレーションを行い、提案手法の有用性を示す。提案手法では一つないし複数の独自因子から0の位置に関して知見がある場合には式(24)（もしくは式(25)）の $r_{k_2y}$ を-1~1の範囲で適切に定めることにより、ある偏回帰係数を0にするか0に最も近づけることができる。本シミュレーションでは、この際の推定精度を検証する。

本シミュレーションでは、ある単一変数の独自因子からは目的変数に対する偏回帰係数が0であることが判明している場合における変回帰係数の推定値の再現度を検証する。比較対象として、共通因子得点と当該の説明変数以外の説明変数を用いて回帰分析を行なったものを用いた。つまり偏回帰係数が0であると判明している変数を削減して分析したものを比較対象とした。共通因子得点の推定方法としては回帰法を用いた。これは、図(1)からもわかるように、シミュレーション1のシチュエーションにおいては Bartlett 法よりも回帰法の方が総じて真値の再現性が高かったためである。また比較対象の回帰分析においては、シミュレーション1のような説明変数間の完全な多重共線性はないものの強い多重共線性が生じているため、式(2)で示される最小二乗推定の他に Lasso 法を用いた偏回帰係数の推定も行なった。

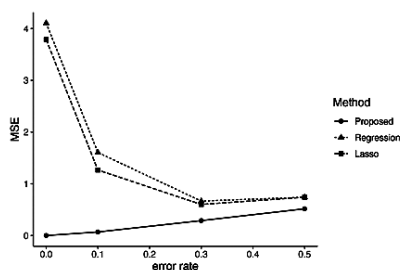


図7 推定手法別 MSE

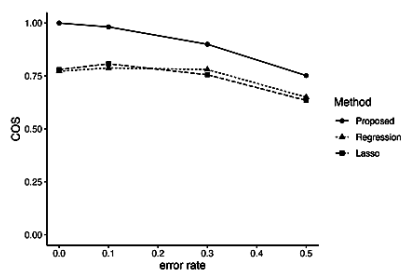


図8 推定手法別 COS

シミュレーションのデータの生成や再現性の評価方法についてはシミュ

レーション1と基本的には同一である。ただし、条件についてサンプルサイズと説明変数の数はサンプルサイズは  $n=100$ 、説明変数の数は  $p=5$  と固定した。また、真の偏回帰係数のうち一つの係数は0に定め、その位置は既知であるとした。提案手法においては式(25)をもとに当該の偏回帰係数が0に最も近づくように求め、比較手法においては当該の変数を削減して推定を行った。提案手法と比較手法のそれぞれについて、推定値と真値のMSEおよびCOSを計算し、その100データセットにおける平均値を当該の条件の真値の再現性とした。

図7および図8は各条件における各手法のMSEおよびCOSの平均値を示したものである。なお、エラーレートは $\alpha_x = \alpha_y$ である。図から示されているように、まずMSEはどのエラーレートの水準においても提案手法の値が小さい。また、COSの値もいずれの水準においても提案手法の値が大きい。これらのことは、MSEとCOSのどちらの観点からも提案手法の真値の再現性が比較手法に比べて高いことを示している。

## 実データを用いた解析例

本章では提案手法の分析結果の妥当性を示すために、実データを用いた解析例を示す。分析対象とするデータはHolzinger and Swineford(1939)の24 psychological test dataである。このデータセットはHarman(1976)などでも利用されており、因子分析の分析対象として代表的なデータセットである。また、本分析で用いたデータセットはRのパッケージpsychTools(Revelle, 2024)中で“holzinger.swineford”として提供されているものを用いた。データセットは24項目の心理テスト（能力テスト）で構成されているが、本分析ではこのうち言語能力を測定している以下の5項目

- t05\_geninfo: General Information Verbal Test（一般情報の口述）
- t06\_paracomp: Paragraph Comprehension Test（パラグラフの理解）
- t07\_sentcomp: Sentence Completion Test（文の穴埋め）
- t08\_wordclas: Word classification-Which word not belong in set（単語の分類）
- t09\_wordmean: Word Meaning Test（単語の語義）

を用いた。これらの項目は言語能力を測定するために用いられたテスト項目であり、実際にこれらの項目に対して平行分析などで因子数を判断すると1因子モデルが支持される。

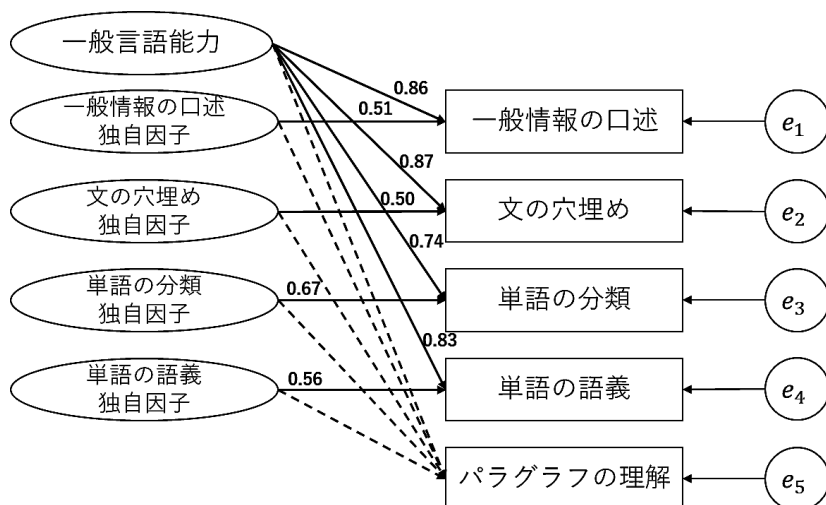


図9 分析モデル図

本分析では、パラグラフの理解が他の変数の独自因子項目および言語能力因子から影響を受けているというモデルのもとで分析を行った。つまり、モデルの概要図は図1のように描ける。なお、図1の破線は提案手法で推定対象としているパラメータを示している。またここで、 $e_1, \dots, e_5$  は誤差項である。なお実線部の数値は因子負荷を示しており、これらは説明変数に対して因子数1の因子分析を行って得られたものである。図9からも分かるように、一般言語能力因子の統制を行ったうえでのそれぞれの顕在変数に独自に影響を与えている独自因子がパラグラフの理解にどれだけ影響を与えているかを推定する。

分析の結果得られた偏回帰係数は表2の通りである。推定結果の妥当性の検証のために、Regression法を用いて求めた因子得点と顕在変数を同時

に説明変数として用いた重回帰モデルの偏回帰係数をRidge法およびLasso法を用いて推定した結果および共通因子を利用しない回帰分析の結果も示されている。また、提案手法の推定結果の括弧内の数値は95%信頼区間の下端と上端を示している。表からもわかるように、「一般言語能力」因子を統制したもつで、「一般情報の口述」および「単語の分類」は「パラグラフの理解」に与える影響は限定的であると解釈できる点において、提案手法と比較手法で結果は概ね一致している。このことから、提案手法の分析結果は比較手法と著しい乖離を生じておらず、分析結果の解釈における全体的な妥当性が示唆されている。一方で、提案手法と比較手法の結果における差異は、「一般言語能力」因子の偏回帰係数の値である。数値シミュレーションの結果を踏まえて考慮すると、図1のモデルが妥当であれば、比較手法では共通因子が与える影響の大きさを過小推定している一方で、提案手法では適切に推定されていると考えられる。また、結果の解釈の面から判断しても、Ridge法で得られた偏回帰係数では一般言語能力因子がパラグラフの理解に与える影響は文の穴埋めよりも小さいとされているが、これは考えづらいだろう。以上のように、提案手法での分析結果には一定の妥当性がみられた。

表2 偏回帰係数の推定値（括弧内は信頼区間）

	提案手法	Ridge	Lasso	回帰分析
一般言語能力	0.74(0.66, 0.81)	0.21	0.40	—
一般情報の口述	0.05(-0.02, 0.13)	0.07	0	0.11
文の穴埋め	0.19(0.12, 0.26)	0.28	0.23	0.37
単語の分類	0.05(-0.02, 0.13)	0.06	0	0.08
単語の語義	0.17(0.10, 0.24)	0.23	0.18	0.30

また、提案手法における区間推定を利用して有意性の判断を行うと、「一般言語能力」「文の穴埋め」「単語の語義」の偏回帰係数が5%信頼係数で有意になっているが、「一般情報の口述」「単語の分類」は統計的に有意ではない。このことは「一般情報の口述」や「単語の分類」の独自因子は「パラグラフの理解」に影響を及ぼしていない可能性があることを

示している。例えば「一般言語能力」と関連がなく「一般情報の口述」の得点に独自に影響する要素としては応答の際の口調の巧拙などが考えられ、「単語の分類」に独自に影響する要素としてはドメイン知識などが考えられる。提案手法による分析結果によれば、それらは「パラグラフの理解」に影響を及ぼさない可能性を否定できないということを示しており、この結果は納得できるだろう。

## 考察

本論文では、重回帰分析において説明変数間に強い相関関係があるような場合に、潜在変数を共通因子と独自因子として再構造化し、それらを説明変数として用いた重回帰分析の提案を行った。これにより、共通因子と独自因子との影響の違いがより明らかになり、データの解釈に有用な知見を与えることを目的としている。因子分析には因子得点の不定性が存在しているため、本研究では因子得点の不定部分を「説明変数とは無関係であるが、共通因子・独自因子とは関連している情報と目的変数との相関関係の強さ」を表す相関係数として解釈することにより、一意な偏回帰係数を推定した。また、数値シミュレーションの結果、提案手法で推定された偏回帰係数は比較手法よりもより真値の再現性の高い推定を行うことができることが明らかになった。さらに、実データ解析の結果から、提案手法の分析結果の妥当性が示唆された。

本研究では、因子分析における因子得点の不定性について、回帰分析を行う際の目的変数との相関係数という観点で整理を行ったが、現実場面においては因子得点が不定であることは問題となり得る。このことは、現実の問題を対処するには回帰係数の信頼性という形で問題として現れる。しかしながら、本研究における、因子得点を利用した際の推定値の取り得る値の範囲という考え方はその問題を取り扱う上での一助になると考えられる。例えばシミュレーションでも示したような、「最も悪い状況でもこの範囲に含まれる」という知見は検証的な調査を設計する上で有効に活用できるだろう。

また、提案手法は因果推論への応用も考えらえる。観察データを用いた重回帰分析において、しばしば問題となるのは、複数の説明変数に関わる未観測の交絡因子の存在である。本研究で得られた、説明変数に因子分析モデルをあてはめることが適切な状況では、未観測の交絡因子が説明変数に与える影響を因子負荷行列として与えることによって、観察データの発生源である独自因子が目的変数に与える影響を推定する、感度分析が行える可能性がある。また、先述した推定値が取り得る値の範囲を利用することによって、交絡因子が与える影響を定量的に考察することができるかもしれない。同様に、その他の社会科学においても図9のようなモデル図で表されるような状況は大いにありえるだろう。一方で、因果推論の分野においては、処置などの効果や実際に観測できる変数の効果を推定することが多い。そのため、本研究で取り扱ったような観測された説明変数群を因子分析モデルを用いて再構造化することが適当である状況、つまりは観測された説明変数群の変動は複数の変数に影響を与える共通因子の変動と単一の変数のみに与える独自因子の変動で説明できるとみなせ、それらの因子の影響に興味がある状況、は興味の対象外であることが多いかもしれない。本研究で取り扱った状況では提案手法は妥当な結論を導くことができると考えているが、因果推論などへの応用可能性の検証については今後の研究が望まれる。

## 謝辞

本研究はJSPS科研費JP22H03681およびJSPS科研費JP21K13393の助成を受けたものです。

## 参考文献

- Adachi, K. (2012). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis, *Journal of the Japanese Society of Computational Statistics*, 25, 25-38.
- Akaike, H. (1973). Information theory and an extension of the maximum

- likelihood principle, Proceedings of the 2nd International Symposium on Information Theory, Budapest, 267-281.
- Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28, 97-104.
- Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, no. 48. Chicago: University of Chicago, Department of Education.
- Harman, H. H. (1967). *Modern factor analysis*. Chicago, University of Chicago Press.
- Hoerl, A. E., & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, 1970, 12, 55-67.
- 岩崎 学 (2021) 統計的因果推論の視点による重回帰分析, *日本統計学会誌*, 2021, 50, 363-379.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis (M&T series)*, DSWO Press, Leiden University.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 58, No. 1, 267-288.
- Unkel, S., & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review, *International Statistical Review*, Vol. 78, No. 3, 363-382.
- Uno, K., Adachi, K., & Trendafilov, N. T. (2019). Clustered Common Factor Exploration in Factor Analysis, *Psychometrika*, Vol. 84, No. 4, 1048-1067.
- Venables, W. N., & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York.
- Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois.



Revelle, W. (2024). psychTools: Tools to Accompany the 'psych' Package for Psychological Research. Northwestern University, Evanston, Illinois.

吉田 寿夫 & 村井 潤一郎 (2021). 心理学的研究における重回帰分析の適用に関わる諸問題. 心理学研究, 92(3), 178-187.