

博士論文

順序カテゴリデータから計算される相関係数に  
生じるバイアスおよびその対処法の評価

2021年度

小野島 昂洋

早稲田大学

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	社会科学・行動科学における順序カテゴリデータ . . . . .	1
1.2	順序カテゴリデータから計算される相関係数に生じるバイアス . . . . .	3
1.3	相関係数のバイアスが心理学研究にもたらす問題 . . . . .	5
1.4	本研究の目的および構成 . . . . .	6
<b>第 2 章</b>	<b>順序カテゴリデータから計算される相関係数の数理</b>	<b>8</b>
2.1	カテゴリ化後の変数から計算される相関係数 . . . . .	8
2.2	カテゴリ化による相関係数のバイアスに閾値が及ぼす影響 . . . . .	16
2.3	本章のまとめ . . . . .	19
<b>第 3 章</b>	<b>カテゴリ化によって生じる相関係数のバイアスの評価</b>	<b>20</b>
3.1	カテゴリ化によるバイアスを評価した先行研究 . . . . .	20
3.2	離散化の際に閾値が多様な値をとりうる場合のバイアスの評価 (研究 1) . . . . .	24
3.3	本章のまとめ . . . . .	30
<b>第 4 章</b>	<b>カテゴリ化によって生じる相関係数のバイアスへの対処法の評価</b>	<b>32</b>
4.1	カテゴリ化によるバイアスへの対処法の先行研究 . . . . .	32
4.2	カテゴリ値に整数値以外を割り振った場合および polychoric 相関 係数の比較 (研究 2-a) . . . . .	38
4.3	カテゴリ値に整数値以外を割り振った場合および polychoric 相関 係数の実データによる検討 (研究 2-b) . . . . .	53

4.4	本章のまとめ . . . . .	57
<b>第 5 章</b>	<b>カテゴリ化が信頼性係数に与える影響</b>	<b>58</b>
5.1	信頼性係数の利用についての先行研究 . . . . .	58
5.2	連続変数のカテゴリ化が Cronbach の $\alpha$ へと及ぼす影響 (研究 3) .	67
5.3	順序カテゴリデータに対する因子分析に基づく信頼性係数の評価 (研究 4-a) . . . . .	74
5.4	順序カテゴリデータに対する因子分析に基づく信頼性係数の実デー タへの適用 (研究 4-b) . . . . .	92
5.5	本章のまとめ . . . . .	96
<b>第 6 章</b>	<b>結論と展望</b>	<b>97</b>
6.1	各章のまとめ . . . . .	97
6.2	本研究の意義 . . . . .	98
6.3	本研究の課題と今後の展望 . . . . .	101
	<b>参考文献</b>	<b>103</b>
	<b>付録 A 研究 1 の補足資料</b>	<b>115</b>
	<b>付録 B 研究 2-a の補足資料</b>	<b>128</b>
	<b>付録 C 研究 3 の補足資料</b>	<b>139</b>
	<b>付録 D 研究 4-a の補足資料</b>	<b>143</b>
	<b>初出一覧</b>	<b>157</b>

# 第 1 章

## 序論

### 1.1 社会科学・行動科学における順序カテゴリデータ

心理学を含む社会科学・行動科学にとって「測定」は極めて重要な問題である。Stevens (1946) は測定を「ルールに従って対象や事象に数値を割り当てること (p.677)」と定義したが、この定義は心理学において広く知られ心理測定のテキストにおいてもしばしば紹介されている (Guilford, 1954; 田中, 1977)。対象へ数値を割り当てる際のルールには様々なものが考えられ、異なる数値の割り当て方からは異なる種類の尺度が構成される。Stevens (1946) は、対象を測定する際の経験的な操作、その結果として生じた尺度の持つ数学的性質、尺度から許容される統計量の 3 つの観点から「名義」「順序」「間隔」「比例」の 4 つの尺度水準を提案し、この分類は多くの心理測定のテキストにおいて採用されている (e.g., Guilford, 1954; 田中, 1977; Furr, 2018)<sup>1</sup>。

Stevens の 4 つの尺度水準では「名義」「順序」「間隔」「比例」の順番に適用可能な統計的操作の種類が多くなる。例えば、名義尺度において許容される統計量はケースの数やモードであるが、順序尺度ではそれに加えて中央値やパーセンタイルが、間隔尺度では平均や標準偏差なども意味ある統計量として算出できる。物理学などの

---

<sup>1</sup> Stevens (1946) による尺度水準の他にも、4 つの尺度の様々な変異を検討したもの (Coombs, Raiffa, & Thrall, 1954) や、Stevens の「名義」「順序」尺度は、測定というよりも数字による符号化 (numerical coding) だとする立場 (Michell, 1986) もある。

自然科学においては、長さや重さなどの比例尺度に基づくデータが得られることが多いが、学問領域によっては必ずしも比例尺度が容易に構成できる訳ではない。実際 4 つの尺度水準を提案した Stevens も、心理学において比例尺度が得られることはまれであり、もっとも広く使われているのは順序尺度であると述べている (Stevens, 1946)。

人を対象とする研究の中でも、人の能力や態度などの心理学的属性 (psychological attribute) が測定の対象となる場合に、順序尺度は古くから用いられている。評定尺度を心理学の研究に導入した Galton<sup>2</sup>は人の能力の測定の中で順序尺度を複数用いている。その中でも Galton (1883) によるものは有名で、質問紙を用いて心像 (mental imagery) —— 目の前にないものを心の中で思い浮かべた際にどの程度はっきりと思い浮かべることができるか —— を測定しており、心像の明瞭さや色について 9 段階の順序尺度を作成している。

もちろん現在においても順序カテゴリデータは頻繁に用いられている。その代表的なものとして、態度を測定する Likert 型の項目 (Likert type item) への回答のデータが挙げられる。Likert 型の項目では多くの場合、回答者は提示された文について同意もしくは非同意の程度を 2-7 段階程度の予め決められた選択肢の中から選んで回答し、各カテゴリには通常 1 から始まる整数値が割り当てられる。相互に関連するまとまりのある複数の Likert 型の項目得点を合計 (あるいは平均) して尺度得点として用いることまで含める場合には Likert 尺度 (Likert scale) と呼ばれ、Likert 型の項目と区別されることもある。Likert 尺度が測定の水準において順序尺度か間隔尺度かという論争は昔から存在しているが、Likert 型の項目は順序尺度として扱うのが適当である一方で、Likert 尺度は実用上は間隔尺度として扱えるという主張が増えてきている (Carifio & Perla, 2008; Norman, 2010)。

Likert (1932) は、当時の尺度作成の手続き<sup>3</sup>は多数の尺度項目候補の用意や複数の

---

<sup>2</sup> 後の研究へ影響を与えたという観点から言えば Galton が心理学の研究に評定尺度を導入したと言っても間違いはないと思われるが、Galton を評定尺度の最初の使用者として考えるのは間違っている。Ellson & Ellson (1953) や McReynolds & Ludwig (1987) は Galton 以前に評定尺度の使用例があったことを報告している。

<sup>3</sup> 等間隔法 (method of equal-appearing intervals) や Thurstone 法と呼ばれる手法 (Thurstone, 1928) のことである。手続きの詳細は Guilford (1954) や岡本 (1991) を参照されたい。

評定者の確保など、作成に非常に時間や労力がかかるものだったことを背景に、より簡便に態度の測定を可能にするものを模索し Likert 尺度を開発した (Edmondson, 2005)。Likert 尺度は、広い範囲の構成概念の測定に使えること、紙やウェブでの調査に適した視覚的なフォーマットを持っていること、実施が簡便で迅速であることなどの調査実施に際しての実践的な利点があるため (Revilla, Saris, & Krosnick, 2013)、現在まで継続して用いられている。

Likert 尺度が提案されたのは心理学においてであるが、Likert 尺度および Likert 型の項目を用いた調査は現在までに様々な領域で用いられている。例えば、マーケティング分野における Likert 尺度の使用を調べた Edmondson (2005) は、*Journal of Marketing* 誌上において実証研究の発表数が増えていく 1970 年以降に Likert 尺度の利用が増えていったことを報告している。また、Clason & Dormody (1994) は農業教育の分野における Likert 型項目の使用状況を調べ、ある一定の期間において発表された論文のうちの半分以上を越える論文において Likert 型の項目を用いた測定が行われていることを報告している。医学の分野においては、アウトカムの評価に患者報告式の評定尺度が使われており、その科学的な厳密性を高めることがアメリカ医薬食品局 (FDA) や欧州医薬品庁 (EMA) から求められている (Hobart, Cano, Zajicek, & Thompson, 2007)。

これらの例から分かるように人間を対象とした様々な研究領域において順序尺度を用いて順序カテゴリデータが得られている。しかしながら、次節で見るように順序カテゴリデータから計算される相関係数には絶対値が減少するバイアスが生じることが知られている。

## 1.2 順序カテゴリデータから計算される相関係数に生じるバイアス

社会科学や行動科学の様々な分野で、しばしば変数間の関係が中心的な問題となる。変数間の (線形) 関係を示す統計量として現在までもっとも使われているのは

Pearson の相関係数 (Pearson's correlation coefficient) であろう<sup>4</sup>。相関係数は、天文学や測地学などを具体的な文脈とする誤差法則の研究から生まれ、Galton によって遺伝法則として解釈され、Pearson (1896) によって定義された (椎名, 2016)。

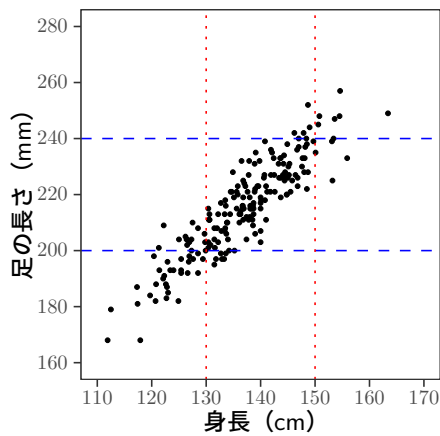
相関係数を順序カテゴリデータから計算した場合に生じる問題の一つに、相関係数の絶対値が減少するバイアスが生じることが知られている。この問題は、相関係数を提案した Karl Pearson 自身も「幅広いカテゴリ (broad category)」の問題としてとりあげておりその対応策も検討されたが (Pearson, 1913; 椎名, 2017), 100 年以上経った今現在もこの問題は完全な解決には至っていない。

ここで順序カテゴリデータから計算される相関係数のバイアスがどのようなものかの概要を示す。例えば、ある小学校の男子児童 200 人分の身長と足の長さのデータが得られているものとする<sup>5</sup>。ここで身長について、130cm より小さい児童を「低い」、130cm から 150cm の間の児童を「普通」、150cm より大きい児童を「高い」と 3 つのカテゴリに分けて、それぞれに 1, 2, 3 とカテゴリ値を割り当てたとする。足の長さについても同様に、200mm 以下を「小さい」、200mm から 240mm を「普通」、240mm より大きい児童を「大きい」と 3 つのカテゴリに分けて、それぞれに 1, 2, 3 とカテゴリ値を割り当てたとする。このように連続変量を一連の閾値 (threshold) でカテゴリ化した場合に、データは分割表 (contingency table) の形で得られることとなる (Figure 1.1)。この分割表のデータから相関係数を計算した場合に相関係数の絶対値は減少することが知られている (Peters & Van Voorhis, 1940)。例えば、Figure 1.1 (a) の散布図は、カテゴリ化される前の身長と足の長さを示しているがそこから標本相関係数を算出すると  $r = .901$  である一方で、Figure 1.1 (b) の分割表で示されるカテゴリ化後の身長と足の長さから計算した相関係数は  $r = .688$  であり、絶対値の低下が見られる。

---

<sup>4</sup> 相関係数には順序相関係数などもあるが、以下では単に相関係数と書いた場合には Pearson の相関係数を指すこととする。

<sup>5</sup> ここで例示に用いたデータは R のパッケージ `lgrdata version 0.1.1` (Duursma, 2019) に所収のデータ (anthropometry) を加工して作成したものである。具体的には 7 歳以上 12 歳以下の男児のデータで欠損値のない先頭 200 ケースを抽出したものをを用いた。



(a) カテゴリ化前 ( $r = .901$ )

n = 200		身長		
		1	2	3
足の長さ	3	0	4	6
	2	13	132	5
	1	32	8	0

(b) カテゴリ化後 ( $r = .688$ )

Figure 1.1. 相関係数のカテゴリ化によって生じるバイアス

### 1.3 相関係数のバイアスが心理学研究にもたらす問題

前節で述べた順序カテゴリデータから計算される相関係数に生じるバイアスは、人間の態度や意識、性格特性などの心理学的属性を対象とする研究領域において特に重要な課題である。なぜなら、研究が対象とするものが実際には観測不能な潜在変数であり、得られる観測データは一連の閾値によってカテゴリ化がなされた後のデータであるケースが多いからである。先に態度の研究などにおいて Likert 型の項目が頻繁に用いられることを述べたが、Likert 型の項目は閾値による離散化の前提のもとに利用されていることが多い。心理学など潜在変数を対象とする学問領域において変数間の関係を記述する際に相関係数は主要な道具であるが、その相関係数にバイアスが生じているのであれば、変数間の関係を正確に記述できないことにつながる。

また、相関係数は変数間の線形関係の強さを記述するのみでなく、相関係数を基にして、信頼性係数の算出が行われたり、回帰分析、因子分析、構造方程式モデリング (Structural Equation Modeling, SEM) などの更なるデータ解析に用いられたりもする。当然のことながら、相関係数において生じたバイアスは回帰分析や因子分



析などのその後の分析の結果を歪めることにつながる。もともと連続変数であったものをカテゴリ化した際に生じる問題については現在まで様々なものが指摘されており、特に変数が2値化された場合についてはよく報告されている。例えば Cohen (1983) は、連続変数を2値化した場合に生じる検出力の低下の程度は、サンプルの1/3から2/3を捨てているのと同程度だと指摘している。また、2値化した変数を用いて分散分析を行う際には、擬似主効果や擬似交互作用が生じることも報告されている (MacCallum, Zhang, Preacher, & Rucker, 2002)。これらの問題は相関係数に生じたバイアスによって引き起こされるものであるため、カテゴリ数が2より大きい場合であっても相関係数に大きなバイアスが生じているのであれば、同様の問題を生じさせることにつながるだろう。また、順序カテゴリデータから因子分析を行う際には、誤った因子を抽出したり、バイアスのかかった因子負荷量が推定されるなどの問題点が指摘されている (Olsson, 1979b; 萩生田・繁樹, 1996)。

注意しなければならないのは、バイアスが存在するからといって必ずしも順序カテゴリデータから相関係数を計算することが不適切とは言えないことである。なぜなら、カテゴリ化によって生じるバイアスが無視できるほど小さいのであれば、相関係数を算出してデータの線形関係を簡潔に記述できることの方がメリットが大きいからである。したがって、バイアスの程度を正確に評価することが重要な課題となる。同様に、相関係数にバイアスが生じるからといって、相関係数を利用したその後の分析を即座にやめなければならないとするのも間違いであろう。肝要なのは、相関係数のバイアスがその後の分析結果に与える影響の程度を評価することである。

バイアスの程度が大きいのであれば、バイアスに対する対処法を検討する必要があるだろう。バイアスへの適切な対処法を選択することで、従来の分析法では見落とされていた（あるいは歪められていた）潜在変数間の関係を発見することにつながる可能性があるからである。

## 1.4 本研究の目的および構成

以上を背景にして本研究は、順序カテゴリデータから計算される相関係数のバイアスを評価し、バイアスに影響を及ぼす要因を検討するとともに、バイアスへの対

処法の評価を行うことを目的とする。また、相関係数そのもののバイアスの評価に加えて、バイアスがその後の分析においてどのような影響を及ぼすかの検討も行う。相関係数を用いる分析には様々なものがあり、その全てを扱うことは本論文の射程を大きく超えるので、本研究では相関係数から算出することができる係数の一つである信頼性係数に焦点を当てる。測定の信頼性はその後のデータ解析が信頼に足るものかを左右する重要な要因であり、信頼性係数の問題を検討することの意義は大きいだろう。

まず、2章ではカテゴリ化された変数から計算される相関係数の数理的な検討を行う。具体的には、カテゴリ化された2変量の期待値、分散、共分散を整理し、カテゴリ化後の相関係数の期待値を検討する。また、変数のカテゴリ化の仕方によってバイアスの程度がどのように変化するかを検討する。続く3章では、相関係数のバイアスの評価を行う。まず、現在までに相関係数のバイアスの程度を評価した研究を概観し先行研究のシミュレーションでは2変数のカテゴリ数が異なる場合や、変数の離散化の際の閾値を十分に考慮していないことを指摘した上で、新たにこれらの要因を考慮したモンテカルロ・シミュレーションを行い、より現実のデータ収集に即したバイアスの程度を評価する（研究1）。4章ではバイアスへの対処法を検討する。バイアスへの対処法としては、polychoric 相関係数の利用や、Likert 型の項目へ割り当てる数値を工夫する方法が考えられる。それらの手法を用いた場合に、相関係数のバイアスがどのように低減するかをモンテカルロ・シミュレーションによって検討する（研究2-a）。また、実データに対してこれらの手法を適用した場合の例示も行う（研究2-b）。5章では、相関係数のバイアスが信頼性係数へと与える影響を検討する。心理学における信頼性係数の利用の現状をまとめたのちに、信頼性係数の内現在もっとも広く使われている Cronbach の  $\alpha$  へと相関係数のバイアスが与える影響を検討するモンテカルロ・シミュレーションを行う（研究3）。また、順序カテゴリデータ向けの信頼性として近年提案された非線形 SEM 係数（nonlinear SEM coefficient）の性能評価を行うモンテカルロ・シミュレーションを行い（研究4-a）、その実データへの適用も試みる（研究4-b）。最後の6章では各章をまとめ、残された課題を整理して今後の展望を述べる。

## 第 2 章

# 順序カテゴリデータから計算される相関係数の数理

本章では、順序カテゴリデータから計算される相関係数の数理的な検討を行う。2.1 節では、2 変量正規分布から生成されたデータが一連の閾値によってカテゴリ化された際の期待値・分散・共分散を求めることによってカテゴリ化後の相関係数の期待値を導出する。続く 2.2 節では、カテゴリ化の際の閾値がカテゴリ化による相関係数のバイアスへと及ぼす影響を検討する。

### 2.1 カテゴリ化後の変数から計算される相関係数

#### 2.1.1 閾値モデルによるデータ生成の前提

2 次元の確率変数ベクトル  ${}^t(U, V)$  が、期待値ベクトル  $\boldsymbol{\mu}$  と分散共分散行列  $\Sigma$  をパラメータとする 2 変量正規分布に従うものとする。以後の議論において

$$\boldsymbol{\mu} = \mathbf{0}, \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (2.1)$$

とにおいても一般性は失われない。このとき  ${}^t(U, V)$  が従う分布の同時密度関数は

$$\phi(u, v | \boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{(u^2 - 2\rho uv + v^2)}{2(1-\rho^2)}\right] \quad (2.2)$$

である。このとき  $U$  と  $V$  がそれぞれ  $m - 1$  個からなる閾値  $\theta_i$  および  $n - 1$  個からなる閾値  $\tau_j$  でカテゴリ化されることにより  $X$  と  $Y$  が観測される閾値モデル (threshold model) を考える。すなわち,

$$x = \begin{cases} 1, & \text{if } u \leq \theta_1 \\ 2, & \text{if } \theta_1 < u \leq \theta_2 \\ \vdots & \vdots \\ m - 1, & \text{if } \theta_{m-2} < u \leq \theta_{m-1} \\ m, & \text{if } \theta_{m-1} < u \end{cases}, \quad y = \begin{cases} 1, & \text{if } v \leq \tau_1 \\ 2, & \text{if } \tau_1 < v \leq \tau_2 \\ \vdots & \vdots \\ n - 1, & \text{if } \tau_{n-2} < v \leq \tau_{n-1} \\ n, & \text{if } \tau_{n-1} < v \end{cases} \quad (2.3)$$

である (Figure 2.1)。本節では、カテゴリ化後の変数  $X$  と  $Y$  の性質を考える。

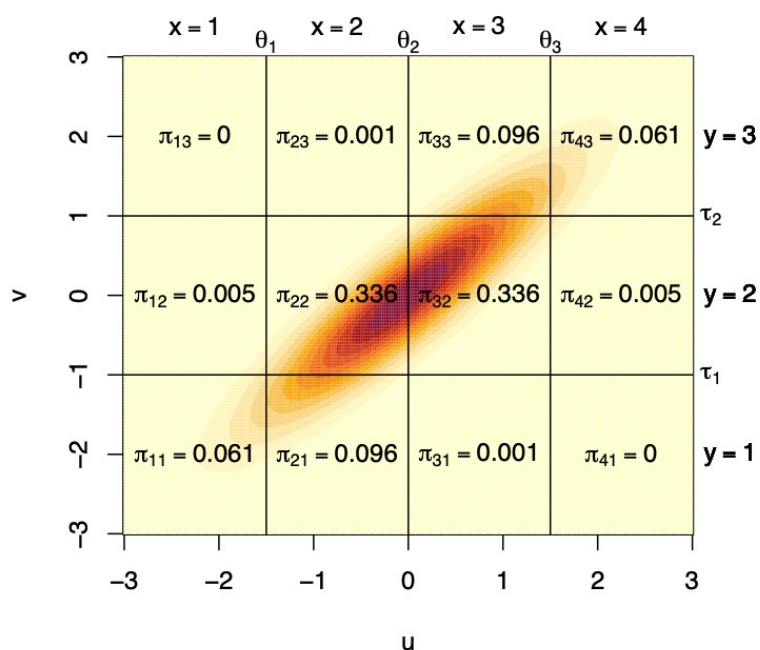


Figure 2.1. 閾値モデルによる変数のカテゴリ化 ( $\pi_{ij}$  は、 $U$  と  $V$  が各セルに落ちる確率)

## 2.1.2 カテゴリ化された確率変数の期待値・分散・共分散

離散型の分布であるカテゴリ変数  $X, Y$  の同時確率関数  $p(i, j) = P[(X = i) \wedge (Y = j)]$  は,

$$p(i, j) = \int_{\theta_{i-1}}^{\theta_i} \int_{\tau_{j-1}}^{\tau_j} \phi(u, v) dv du \quad (2.4)$$

である。ただしここで  $\theta_0 = \tau_0 = -\infty, \theta_m = \tau_n = \infty$  である。以下では、簡単に  $p(i, j) = \pi_{ij}$  と表す。また周辺確率については、それぞれ  $\pi_{i\cdot} = \sum_{j=1}^n \pi_{ij}$ ,  $\pi_{\cdot j} = \sum_{i=1}^m \pi_{ij}$  とする。

$X$  と  $Y$  の相関係数は、それぞれの期待値を  $\mu_X, \mu_Y$ , 分散を  $\sigma_X^2, \sigma_Y^2$ , 共分散を  $\sigma_{XY}$  と表せば,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} \quad (2.5)$$

によって与えられる。

$\mu_X$  と  $\mu_Y$  は期待値の定義に従い,

$$\begin{aligned} \mu_X = E[X] &= \sum_{i=1}^m i \pi_{i\cdot} = \sum_{i=1}^m i \int_{\theta_{i-1}}^{\theta_i} \phi(u) du, \\ \mu_Y = E[Y] &= \sum_{j=1}^n j \pi_{\cdot j} = \sum_{j=1}^n j \int_{\tau_{j-1}}^{\tau_j} \phi(v) dv, \end{aligned} \quad (2.6)$$

である。ここで  $\phi(\cdot)$  は 1 変量標準正規分布の確率密度関数である。

$X$  の分散  $\sigma_X^2 = E[(X - \mu_X)^2]$  と  $Y$  の分散  $\sigma_Y^2 = E[(Y - \mu_Y)^2]$  はそれぞれ公式  $\sigma_Z^2 = E[Z^2] - \mu_Z^2$  より,

$$\begin{aligned} \sigma_X^2 &= \sum_{i=1}^m i^2 \pi_{i\cdot} - \left( \sum_{i=1}^m i \pi_{i\cdot} \right)^2 \\ \sigma_Y^2 &= \sum_{j=1}^n j^2 \pi_{\cdot j} - \left( \sum_{j=1}^n j \pi_{\cdot j} \right)^2 \end{aligned} \quad (2.7)$$

である。また、共分散  $\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$  は、

$$\sigma_{XY} = \sum_{i=1}^m \sum_{j=1}^n (i - \mu_X)(j - \mu_Y)\pi_{ij} \quad (2.8)$$

である。

(2.6)(2.7)(2.8) は、1 変量標準正規分布の累積分布関数  $\Phi_1(\cdot)$  および、2 変量標準正規分布の累積分布関数  $\Phi_2(\cdot, \cdot)$  を用いて表すこともできる。後に 5 章で扱う非線形 SEM 係数の定義式でも、累積分布関数を用いて分散と共分散が示されるためここで検討する。

まず期待値について、 $\int_{\theta_{i-1}}^{\theta_i} \phi(u)du = \Phi_1(\theta_i) - \Phi_1(\theta_{i-1})$  であること、および  $\Phi_1(\theta_m) = \Phi_1(\infty) = 1$  に注意すると (2.6) の  $\mu_X$  は、

$$\begin{aligned} \sum_{i=1}^m i \int_{\theta_{i-1}}^{\theta_i} \phi(u)du &= 1 \cdot \Phi_1(\theta_1) + 2 \cdot [\Phi_1(\theta_2) - \Phi_1(\theta_1)] + \cdots + m \cdot [\Phi_1(\theta_m) - \Phi_1(\theta_{m-1})] \\ &= (1 - 2)\Phi_1(\theta_1) + (2 - 3)\Phi_1(\theta_2) + \cdots + m\Phi_1(\theta_m) \\ &= m - \sum_{i=1}^{m-1} \Phi_1(\theta_i) \end{aligned} \quad (2.9)$$

と書き換えることができる。 $\mu_Y$  も同様に

$$\mu_Y = n - \sum_{j=1}^{n-1} \Phi_1(\tau_j) \quad (2.10)$$

である。

また, (2.7) の  $\sigma_X^2$  についても同様に,

$$\begin{aligned}
& \sum_{i=1}^m i^2 \int_{\theta_{i-1}}^{\theta_i} \phi(u) du \\
&= 1^2 \cdot \Phi_1(\theta_1) + 2^2 \cdot [\Phi_1(\theta_2) - \Phi_1(\theta_1)] + \cdots + m^2 \cdot [\Phi_1(\theta_m) - \Phi_1(\theta_{m-1})] \\
&= (1^2 - 2^2)\Phi_1(\theta_1) + (2^2 - 3^2)\Phi_1(\theta_2) + \cdots + m^2 \cdot \Phi_1(\theta_m) \\
&= m^2 + \sum_{i=1}^{m-1} [i^2 - (i+1)^2]\Phi_1(\theta_i) \\
&= m^2 - \sum_{i=1}^{m-1} (2i+1)\Phi_1(\theta_i) \tag{2.11}
\end{aligned}$$

であることから,

$$\begin{aligned}
\sigma_X^2 &= m^2 - \sum_{i=1}^{m-1} (2i+1)\Phi_1(\theta_i) - \left( m - \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right)^2 \\
&= m^2 - \sum_{i=1}^{m-1} (2i+1)\Phi_1(\theta_i) - \left( m^2 - \sum_{i=1}^{m-1} 2m\Phi_1(\theta_i) + \left[ \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right]^2 \right) \\
&= \sum_{i=1}^{m-1} (2m - 2i - 1)\Phi_1(\theta_i) - \left( \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right)^2 \tag{2.12}
\end{aligned}$$

と表せる。 $\sigma_Y^2$  も同様に,

$$\sigma_Y^2 = \sum_{j=1}^{n-1} (2n - 2j - 1)\Phi_1(\tau_j) - \left( \sum_{j=1}^{n-1} \Phi_1(\tau_j) \right)^2 \tag{2.13}$$

である。

共分散については,  $\pi_{ij} = \Phi_2(\theta_i, \tau_j) - \Phi_2(\theta_{i-1}, \tau_j) - \Phi_2(\theta_i, \tau_{j-1}) + \Phi_2(\theta_{i-1}, \tau_{j-1})$  であることに注意して公式  $\sigma_{XY} = E[XY] - \mu_X\mu_Y$  の右辺第 1 項を考えると

$$\begin{aligned}
E[XY] &= \sum_{i=1}^m \sum_{j=1}^n ij\pi_{ij} \\
&= \sum_{i=1}^m \sum_{j=1}^n ij[\Phi_2(\theta_i, \tau_j) - \Phi_2(\theta_{i-1}, \tau_j) - \Phi_2(\theta_i, \tau_{j-1}) + \Phi_2(\theta_{i-1}, \tau_{j-1})] \tag{2.14}
\end{aligned}$$

と表せる。これを展開したとき  $\Phi_2(\theta_i, \tau_j)$  の係数を  $a_{ij}$  とすると

$$a_{ij} = \begin{cases} mn & \text{if } (i = m) \wedge (j = n) \\ -n & \text{if } (i = m) \wedge (1 \leq j \leq n-1) \\ -m & \text{if } (1 \leq i \leq m-1) \wedge (j = n) \\ 1 & \text{if } (1 \leq i \leq m-1) \wedge (1 \leq j \leq n-1) \end{cases} \quad (2.15)$$

となるので  $\Phi_2(\theta_i, \tau_n) = \Phi_2(\theta_i, \infty) = \Phi_1(\theta_i)$  および  $\Phi_2(\theta_m, \tau_j) = \Phi_2(\infty, \tau_j) = \Phi_1(\tau_j)$  であることに注意すると, (2.14) 式は

$$\begin{aligned} & \sum_{i=1}^m \sum_{j=1}^n ij [\Phi_2(\theta_i, \tau_j) - \Phi_2(\theta_{i-1}, \tau_j) - \Phi_2(\theta_i, \tau_{j-1}) + \Phi_2(\theta_{i-1}, \tau_{j-1})] \\ &= mn\Phi_2(\theta_m, \tau_n) + \sum_{i=1}^{m-1} -n\Phi_2(\theta_i, \tau_n) + \sum_{j=1}^{n-1} -m\Phi_2(\theta_m, \tau_j) + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \Phi_2(\theta_i, \tau_j) \\ &= mn - n \sum_{i=1}^{m-1} \Phi_1(\theta_i) - m \sum_{j=1}^{n-1} \Phi_1(\tau_j) + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \Phi_2(\theta_i, \tau_j) \end{aligned} \quad (2.16)$$

と書き換えることができる。一方で  $\sigma_{XY} = E[XY] - \mu_U\mu_V$  の右辺第 2 項は

$$\begin{aligned} \mu_X\mu_Y &= \left( m - \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right) \left( n - \sum_{j=1}^{n-1} \Phi_1(\tau_j) \right) \\ &= mn - m \sum_{j=1}^{n-1} \Phi_1(\tau_j) - n \sum_{i=1}^{m-1} \Phi_1(\theta_i) + \sum_{x=1}^{m-1} \Phi_1(\theta_x) \sum_{y=1}^{n-1} \Phi_1(\tau_y) \end{aligned} \quad (2.17)$$

である。従って (2.16) と (2.17) を用いると, (2.8) は

$$\sigma_{XY} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \Phi_2(\theta_i, \tau_j) - \left( \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right) \left( \sum_{j=1}^{n-1} \Phi_1(\tau_j) \right) \quad (2.18)$$

と表すことができる。

(2.12)(2.13) (2.18) を用いると, カテゴリ化された 2 変量の相関係数を表す (2.5) は



$\rho_{XY}$

$$= \frac{\sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \Phi_2(\theta_i, \tau_j) - \left( \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right) \left( \sum_{j=1}^{n-1} \Phi_1(\tau_j) \right)}{\sqrt{\sum_{i=1}^{m-1} (2m - 2i - 1) \Phi_1(\theta_i) - \left( \sum_{i=1}^{m-1} \Phi_1(\theta_i) \right)^2} \sqrt{\sum_{j=1}^{n-1} (2n - 2j - 1) \Phi_1(\tau_j) - \left( \sum_{j=1}^{n-1} \Phi_1(\tau_j) \right)^2}} \quad (2.19)$$

と書き換えることができる。

なお、(2.19) が成り立つのは (2.3) のようにカテゴリ値に等間隔の整数を割り当てた場合のみである。例えば、カテゴリ値に  $[1, 2, 4, 7]$  といったような数値を割り当てた場合には成立しない。

### 2.1.3 等距離に分割された閾値から計算される相関係数

(2.19) を見ると分かるように、カテゴリ化後の 2 変量の相関係数は、もとの 2 変量正規分布における母相関係数  $\rho_{UV}$  および閾値  $\theta_i$  と  $\tau_j$  の関数である。

母相関係数を 0 から 1.0 まで、閾値の数を 1 から 6 (すなわちカテゴリ数は 2 から 7) まで変化させながら、(2.19) から相関係数を計算したものが Figure 2.2 である。

ここで、Figure 2.2 (a) の  $\theta_i$  および  $\tau_j$  については  $[-3, 3]$  の区間を等間隔で分割する  $m - 1$  個および  $n - 1$  個の値、Figure 2.2 (b) については  $[-1, 1]$  の区間を等分割する点を用いた。具体的には、(a) の閾値は

$$\begin{aligned} \theta_0 &= -\infty, & \theta_i &= -3 + \frac{6}{m}i \quad (i = 1, 2, m - 1), & \theta_m &= \infty \\ \tau_0 &= -\infty, & \tau_j &= -3 + \frac{6}{n}j \quad (j = 1, 2, n - 1), & \tau_n &= \infty \end{aligned}$$

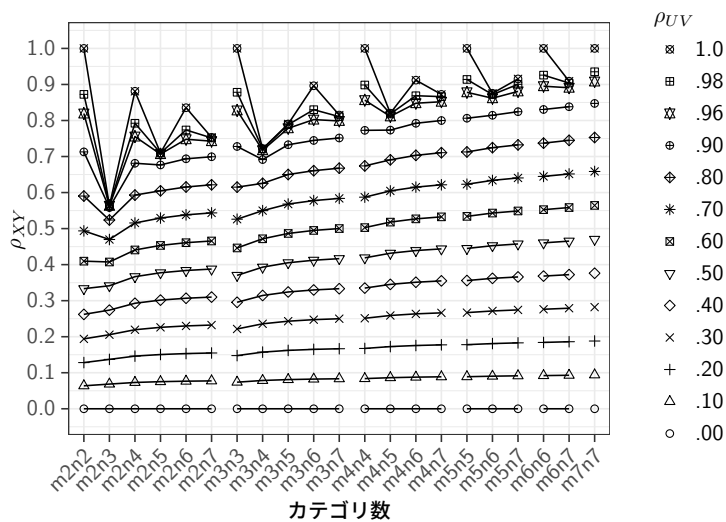
によって定義した。また (b) の閾値は、

$$\begin{aligned} \theta_0 &= -\infty, & \theta_i &= -1 + \frac{2}{m}i \quad (i = 1, 2, m - 1), & \theta_m &= \infty \\ \tau_0 &= -\infty, & \tau_j &= -1 + \frac{2}{n}j \quad (j = 1, 2, n - 1), & \tau_n &= \infty \end{aligned}$$

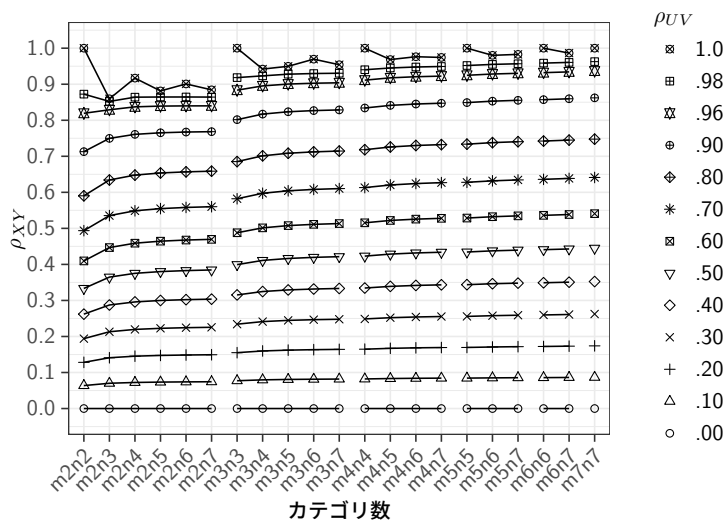
によって定義した。

全体的な傾向として例外<sup>1</sup>はあるものの、より少ないカテゴリ数では相関係数はよ

<sup>1</sup> 例えば、母相関係数が高い場合のカテゴリ数  $m_2n_2$  と  $m_2n_3$  を比べると後者の方が相関係数は大



(a)  $[-3, 3]$  を等間隔で分割する閾値の場合



(b)  $[-1, 1]$  を等間隔で分割する閾値の場合

Figure 2.2. カテゴリ化後の相関係数の期待値

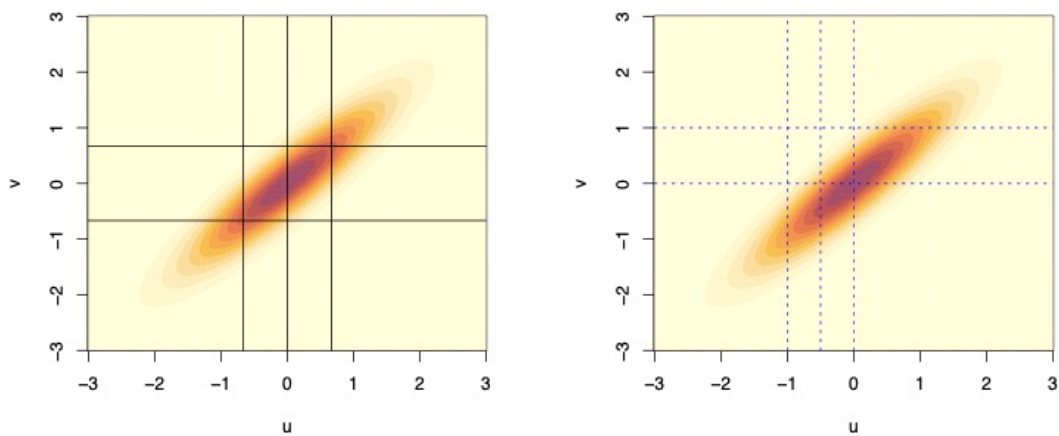
り大きく低下すること、カテゴリ化後の相関係数の減少の程度はもとの2変量正規分布の母相関係数が高いほど大きいこと、カテゴリ化の際に用いられる閾値によってその減少の程度は異なることが見てとれる。

きく低下している。カテゴリ数が異なることによって相関係数が大きく低下する問題は3.1節で検討する。

## 2.2 カテゴリ化による相関係数のバイアスに閾値が及ぼす影響

Figure 2.2 の (a) と (b) を比べれば分かるように、カテゴリ化に伴う相関係数のバイアスの程度は、カテゴリ化の際に用いる閾値によって大きく異なる。本節では、様々な閾値を用いてカテゴリ化した場合に相関係数のバイアスがどのように変化するかを吟味することを通じて、離散化の際の閾値が及ぼす影響を検討する。

Figure 2.3 は 2 変量を同じカテゴリ数ではあるが、異なる閾値でカテゴリ化した例である。カテゴリ化前には  $\rho_{UV} = .900$  あったものが、(a) の閾値（黒の実線）でカテゴリ化した場合には  $\rho_{XY} = .793$  まで低下するが、(b) の閾値（青の点線）でカテゴリ化した場合には  $\rho_{XY} = .662$  まで低下することとなる<sup>2</sup>。



(a) よく整った閾値 ( $\rho_{XY} = .793$ )

(b) 整っていない閾値 ( $\rho_{XY} = .662$ )

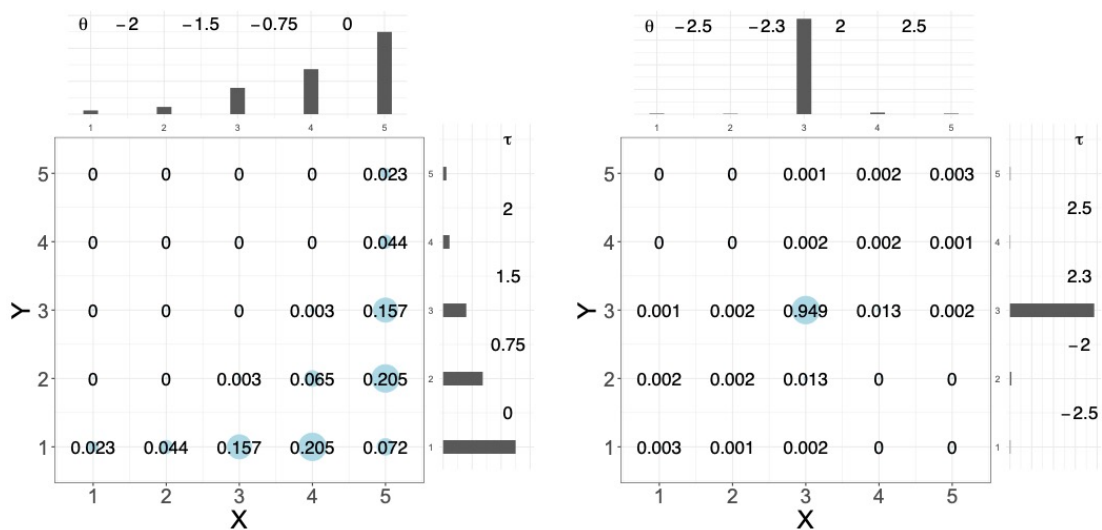
Figure 2.3. 異なる閾値が相関係数のバイアスへと及ぼす影響  
(離散化前  $\rho_{UV} = .900$ )

離散化後の相関係数の低下の度合いが異なる理由は、離散化の際に用いられる閾

<sup>2</sup> 用いられた閾値の具体的な値は (a) では  $\{\theta_1 = -0.6666, \theta_2 = 0, \theta_3 = 0.6666; \tau_1 = -0.6666, \tau_2 = 0.6666\}$ , (b) では  $\{\theta_1 = -1, \theta_2 = -0.5, \theta_3 = 0; \tau_1 = 0, \tau_2 = 1\}$  である。

値によって、もとの連続変数の分布の構造が破壊される度合いが異なることによる。すなわち、ある種の「よく整った (well-organized)」閾値を用いた場合と「整っていない (ill-organized)」閾値を用いた場合では、相関係数の低下の度合いが異なるということである。ここで「よく整った閾値」とは、カテゴリ化に際して、元の2変量正規分布の持つ性質——左右対称、単峰形、分布の中央に空隙がない、1つのセルへの過集中がない——を保つような閾値と定義する。また整っていない閾値はその逆であり、すなわち、非対称、多峰、空隙、1つのセルへの過集中、単調減少や単調増加する分布を生じさせるような閾値と定義する。

これらの性質は質的で曖昧であり、数値的に示すことは難しい。そこで閾値の影響のさらなる理解のために、整っていない閾値によって生み出されるカテゴリカルな分布の例を Figure 2.4 に示した。(a) は、2変量が非対称に歪んでいることにより相関係数が大きく減少する例である。また (b) は、一つのセルへと過集中した場合にはやはり相関係数が大きく減少する例である。どちらの閾値による離散化でも、離散化前の相関  $\rho_{UV} = .900$  から (a) $\rho_{XY} = .584$ , (b) $\rho_{XY} = .595$  と大きく低下している。



(a) 2変量が非対称 ( $\rho_{XY} = .584$ )

(b) 1つのセルへの過集中 ( $\rho_{XY} = .595$ )

Figure 2.4. 整っていない閾値によってカテゴリ化された2変量の分布の例 ( $\rho_{UV} = .900$ , セルの中の値とドットのサイズはセルに落ちる確率)

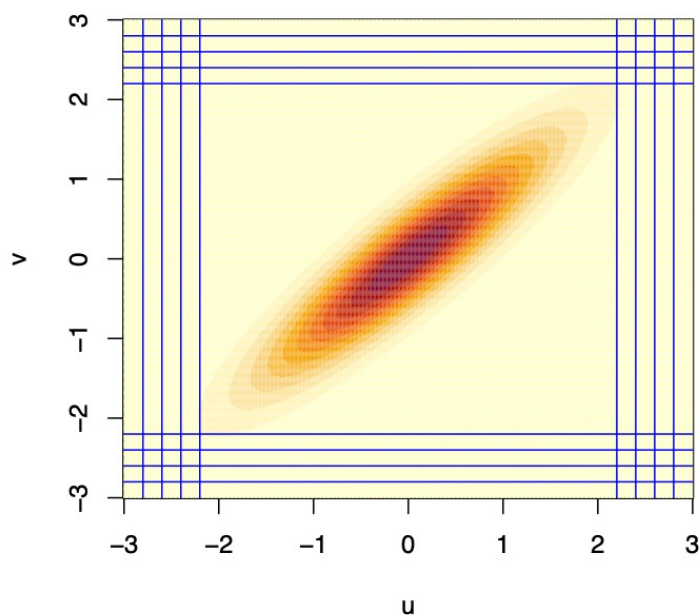


Figure 2.5. 2 変数のカテゴリ数が多いが相関係数が大きく減少する場合  
 $(\rho_{XY} = .657)$

前節の Figure 2.2 について、カテゴリ数が多くなると相関係数のバイアスが小さくなっていく傾向があることを述べたが、閾値の布置によっては必ずしもそうはならないことがある。Figure 2.4(b) のように、一つのセルへと過集中するような閾値を用いてカテゴリ化がされた場合には、カテゴリ数が仮に多かったとしても相関係数のバイアスは大きくなる。Figure 2.5 は  $m = 9, n = 9$  と 2 変数とも比較的多いカテゴリ数であるが、中心のセルへの過集中により離散化前には  $\rho_{UV} = .900$  あったものが、離散化後には  $\rho_{XY} = .657$  まで低下している。

これらの例から分かることは、離散化によって生じる相関係数のバイアスにおいてカテゴリ数は大きな影響を与えることは間違いないのだが、カテゴリ化の際の閾値にも大きく影響を受けるため、バイアスについて検討する際には双方の要因を考慮する必要があるということである。

## 2.3 本章のまとめ

本章では、カテゴリ化された相関係数の数理的性質を検討した。カテゴリ化後の相関係数は、もとの分布の相関係数と閾値の関数であることを示した後、離散化の際の閾値がカテゴリ化によるバイアスへ及ぼす影響を検討した。本章での議論が示唆することは、離散化による相関係数のバイアスはカテゴリ数による影響ももちろん受けるのだが、カテゴリ化の際の閾値にも同様に大きく影響を受けるため、相関係数のバイアスの評価には閾値の影響を含めた検討が必要だということである。

## 第3章

# カテゴリ化によって生じる相関係数のバイアスの評価

前章では、相関係数に生じるバイアスの理解のためにはカテゴリ化の際の閾値を検討する必要があることを指摘した。本章ではその検討を踏まえて、変数のカテゴリ化によって相関係数に生じるバイアスの評価を行う<sup>1</sup>。3.1節では、順序カテゴリデータから計算される相関係数のバイアスを評価した先行研究を概観し、先行研究は2変数のカテゴリ数が異なる場合や変数の離散化の際に用いられる閾値の要因について十分に検討していないことを指摘する。それを受けて、3.2節ではこれらの要因を操作したモンテカルロ・シミュレーション（研究1）を実施し、相関係数のバイアスの程度を評価する。

### 3.1 カテゴリ化によるバイアスを評価した先行研究

#### 3.1.1 社会科学の諸領域における検討

離散化によって生じるバイアスの程度を検討する試みは、初期にはマーケティングサイエンスや政治科学などの社会科学の領域を中心に行われてきた。これはブランドや製品、政策などに対する個人の態度が Likert 型の項目を用いて整数値の離散

---

<sup>1</sup> 本章の内容は Onoshima, Shiina, Ueda, & Kubo (2019) を加筆修正したものである。

データとして得られ、それらが回帰モデルに組み込まれることが多かったためである。Morrison (1972) は、独立変数が従属変数を完璧に予測するモデル（つまり決定係数  $R^2 = 1$ ）から生成されたデータにおいて、従属変数が離散化された場合に決定係数がどの程度減衰するかを検討した。その結果、従属変数が離散的データであることが決定係数に与える影響は小さかったことを報告している。

順序カテゴリ尺度において反応カテゴリの数をいくつにするべきかという長く続く議論を背景に、Martin (1973) は離散的な尺度におけるカテゴリ数の違いが相関係数へと与える影響を検討している。2変量正規分布に従う確率変数を2から20のカテゴリまでカテゴリ数を変えながら離散化した場合に、相関係数がどの程度減少するかの期待値を計算した結果、全体的な傾向としてカテゴリ数が少なくなるほど離散化後の相関係数の絶対値は減少すること、および、もとの2変量の相関係数が高い場合には減少の程度が大きいことを報告し、相関係数はカテゴリカルなデータ取得に対してはロバストでないと結論づけている。

Martin (1978) はこの問題に対してさらに検討を続け、2変量でカテゴリ数が異なる場合に相関係数へと及ぼす影響を調べている。その結果、片方の変数のカテゴリ数が12以上と多い場合には、離散化による相関係数の減少の程度は少ない方のカテゴリ数によって左右されることを報告している。例えば、カテゴリ数の組み合わせが12と4である2変量  $(X_1, Y_1)$  と、組み合わせが16と2である2変量  $(X_2, Y_2)$  では、 $X_1$  と  $X_2$  はともにカテゴリ数が多いので相関係数のバイアスにはほぼ影響を及ぼさず、実質的に  $Y_1$  と  $Y_2$  が相関係数のバイアスが決めるということである。2変量のカテゴリ数の合計は  $(X_1, Y_1)$  では  $12 + 4 = 16$  に対して  $(X_2, Y_2)$  では  $16 + 2 = 18$  なため  $(X_2, Y_2)$  の方が多いが、2つめの変数のカテゴリ数が少ない  $(X_2, Y_2)$  から計算されるバイアスの方が大きくなることが報告されている。

2変量のカテゴリ数が異なる場合の相関係数のバイアスについては大内・上田・椎名・岡田 (2012) も検討している。大内らは (1) 順序カテゴリデータの2変量がそれぞれ異なるカテゴリ数を持ち、(2) それぞれの変数で各カテゴリが最低1回は使われるという条件下では、どのようなデータであっても相関係数は1（あるいは-1）とならないことの証明を与えた後に、モンテカルロ・シミュレーションを行いそのバ



イアスの程度を報告している。その結果、全体的な傾向としてはカテゴリ数が増加すると相関係数のバイアスは減少するが、特定の条件下ではカテゴリ数が増えるともむしろバイアスが大きくなるケースを報告している。具体的には母相関係数が高いとき  $n = m = 2$  と  $n = 2, m = 3$  では、後者の方がカテゴリ数は多いにもかかわらず、相関係数の絶対値がより顕著に低下することを報告している。

Nishisato & Torii (1971) は、2変量の中で片方のみがカテゴリ化された場合と、両方ともカテゴリ化された場合に相関係数に生じるバイアスを検討するモンテカルロ・シミュレーションを行っている。その結果、1変量だけカテゴリ化されるよりも2変量ともにカテゴリ化される方がバイアスが大きいこと、母相関係数が高いほどバイアスが大きいこと、カテゴリ数が少ないほどバイアスが大きいことを報告している。また、Sheppardの補正を用いてこのバイアスが補正できるかも検討したがバイアスを過修正してしまう場合があり上手く機能しないことも報告している。

Bollen & Barb (1981) も相関係数のバイアスに関わる要因——母相関係数、カテゴリ数——を体系的に操作したモンテカルロ・シミュレーションを行っている。2変量正規分布に従う乱数を生成し、連続変数のまま相関係数を計算したものと、カテゴリ化した後に相関係数を計算したものの差を報告している。その結果、カテゴリ化の前後で相関係数に最も差が生じた条件（母相関係数が.8や.9と高く、カテゴリ数が2と少ない場合）であってもその程度は0.2よりも小さく、他の全ての条件ではこれより小さい値であったことを報告している。また、連続変数とカテゴリ化後の相関係数の差は、カテゴリ数の単調関数である（i.e., カテゴリ数が増えるほど差が小さくなる）としている。シミュレーションで報告された相関係数の差が許容可能かどうかについては相関係数を用いる研究の性質や領域によるとしながらも、連続変数の場合とカテゴリ化した場合の相関係数の値の近さをもとに、少なくとも5から6のカテゴリ数が使われるべきであることを主張している。

以上をまとめると、(1) 全体的な傾向としてカテゴリ数が少ない場合に変数の離散化によって大きなバイアスが生じる、(2) 離散化によるバイアスの程度はもとの相関係数の絶対値が大きいほど大きい、ということが一貫している。その一方で、カテゴリ数の増加によってバイアスが低減しないケースがあり、離散化後の相関係数

の値はカテゴリ数の単調関数であるかどうかについては研究によって一貫した結果が得られず結論が出ていないと言える。

### 3.1.2 異なるカテゴリ数の組み合わせおよび閾値の不十分な検討

上述の一貫しない結果を産み出す要因として、2変量のカテゴリ数が異なるケースが十分に検討されていないことが挙げられる。Bollen & Barb (1981) では2変量のカテゴリ数が等しい場合しか扱われていないが、大内他 (2012) では異なる場合が扱われており、この違いがシミュレーションの結果に影響している可能性が考えられる。

また、別の要因としてカテゴリ化の際の閾値の要因が十分に検討されていないことも挙げられる。前章の2.2節で検討したように、カテゴリ化の際の閾値が相関係数のバイアスへと与える影響は大きかったことを考慮すると、この要因を操作することで先行研究間の一貫しない結果の説明が可能だと考えられる。なぜなら、相関係数のバイアスの程度はもとの2変量の構造が破壊される程度によると考えられるが、破壊の程度を実質的に決めているのは閾値の布置だからである。

閾値の要因を検討することは、実際のデータ解析におけるバイアスの程度を把握するのに役にも立つと考えられる。実データの解析において天井効果や床効果などに見られるように、歪んだカテゴリカルな分布が得られることはよくある。これは、もとの連続量の分布が歪んでいるためにこうしたデータが得られたと考えることもできるが、連続量としての心理学的属性がカテゴリ化される際によく整っていない閾値によってカテゴリ化されるために、観測されたデータの分布が歪んでいると解釈することもできる。このことを考慮すると、上述の先行研究 (Nishisato & Torii, 1971; Bollen & Barb, 1981; 大内他, 2012) のシミュレーションで用いられた閾値は、前章の2.2節で検討で検討した閾値のうち「よく整った」閾値に近いので、先行研究で得られた相関係数のバイアスの程度に関する知見は、ある種の理想的な状況下におけるバイアスの程度を評価しているものであると考えるべきである。したがって、現実の多様なデータ収集のプロセスに即したバイアスの程度を評価するためには、離散化の際の閾値の要因を操作したシミュレーションが必要となる。

## 3.2 離散化の際に閾値が多様な値をとりうる場合のバイアスの評価（研究 1）

### 3.2.1 研究の目的

前節の議論において、連続変数の離散化によって生じる相関係数のバイアスを正確に評価するためには、離散化の際の閾値が多様な値をとりうる場合を検討する必要があることを明らかにした。そこで本節ではモンテカルロ・シミュレーションによって、2変数のカテゴリ数が異なる場合および変数のカテゴリ化の際の閾値が多様な場合に相関係数に生じるバイアスの程度を評価することを目的とした。

### 3.2.2 方法

#### 操作した要因と水準

シミュレーションでは (1) 2変量正規分布  $\phi(x, y|0, 0, 1, 1, \rho)$  における母相関係数  $\rho$ , (2) カテゴリ数, (3) サンプルサイズ, (4) 閾値, の4つの要因を操作した。それぞれの要因の水準については、Table 3.1 に示した。

Table 3.1  
研究 1 で操作された要因および水準

操作した要因（水準数）	水準
母相関係数（13）	0, .1, .2, .3, .4, .5, .6, .7, .8, .9, .96, .98, 1.0
カテゴリ数（21）	2 から 7 までのカテゴリ数の全ての組み合わせ
サンプルサイズ（3）	64, 256, 1024
閾値の設定（2）	一様分布条件（10000 通り）、等間隔条件（10000 通り）

母相関係数については、Bollen & Barb (1981) や大内他 (2012) で用いられたのと同じように、小さい値から大きい値までを体系的に操作した。先行研究においては母相関係数が高い場合にとくに相関係数が大きく減少することが指摘されているため、高い母相関係数については細かく水準を設定した。

カテゴリ数については、Likert 型の項目を念頭に質問紙調査で用いられることの

多いカテゴリ数を含むように設定した。サンプルサイズについても、小さいサイズから大きいサイズまでを含むように設定した。

## 閾値の設定

研究の主要な目的の一つが、閾値が相関係数のバイアスへと与える影響を検討することであるため、閾値の要因については2つの条件（水準）を設定した。一つは、連続一様分布から発生させた閾値を用いるもの（一様分布条件）で、もう一つは等間隔の閾値を用いるもの（等間隔条件）である。

$m, n$  を2変量のそれぞれのカテゴリ数としたときに、一様分布条件においては  $m - 1$  および  $n - 1$  個の閾値のペアを連続一様分布  $U(-1, 1)$  から発生させた後に、それぞれを昇順に並べ替えたものを用いた。連続一様分布を用いた理由は、利用可能な事前の情報がない場合に、この分布が一般的に使われるからである。順序統計学を参照すれば、 $l$  を閾値の数としたとき、 $k$  番目の閾値は

$$\text{mean} = -1 + \frac{2k}{l+1}, \quad \text{variance} = \frac{4k(l-k+1)}{(l+1)^2(l+2)} \quad (3.1)$$

をパラメータに持つベータ分布に従うことが知られている (David & Nagaraja, 2003)。したがって閾値の位置は体系的に変動し、中央の位置にある閾値は両端の閾値に比べて大きな分散を持つこととなる。

等間隔条件においては、まず  $[-1, 1]$  の区間をカテゴリ数に等分割する閾値、すなわち  $m$  と  $n$  をそれぞれ閾値の数としたときに、

$$\begin{aligned} \theta_0 &= -\infty, & \theta_i &= -1 + \frac{2}{m}i \quad (i = 1, 2, \dots, m-1), & \theta_m &= \infty \\ \tau_0 &= -\infty, & \tau_j &= -1 + \frac{2}{n}j \quad (j = 1, 2, \dots, n-1), & \tau_n &= \infty \end{aligned}$$

として生成したものに「小さなノイズ<sup>2</sup>」を乗せた閾値を用いた。閾値の順序が入れ替わることを防ぐために、ノイズは切断正規分布から生成した。切断正規分布の確率

<sup>2</sup> カテゴリの判断において各閾値が一定でなく閾値自体も分布するという仮定は、カテゴリ判断の法則 (Law of categorical judgment) として知られ心理測定においては馴染みのあるものである (Torgerson, 1958)。

密度関数は

$$f(z; \mu, \sigma^2, a, b) = \frac{\phi\left(\frac{z-\mu}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right)} \quad (3.2)$$

で与えられる。ここで、 $\phi(\cdot)$  は標準正規分布の密度関数、 $\Phi(\cdot)$  は累積分布関数、 $[a, b]$  は領域（分布の範囲）である。本シミュレーションの文脈では、 $\mu = -1 + 2i/m$ ,  $\sigma^2 = (0.05)^2$ ,  $a = \mu - 1/m$ ,  $b = \mu + 1/m$  と設定することで、閾値  $\theta_i$  の分布は、

$$\theta_i \sim \text{TN}\left(-1 + \frac{2i}{m}, (0.05)^2, -1 + \frac{2i}{m} - \frac{1}{m}, -1 + \frac{2i}{m} + \frac{1}{m}\right) \quad (3.3)$$

で与えられる。この分布においては、 $b - \mu = \mu - a = 1/m$  であるため、 $\theta_i$  の平均、モード、中央値は等しくなる。また、 $-\infty = \theta_0 < \theta_1 < \theta_2 < \dots < \theta_m = \infty$  の制約は常に満たされる。 $\tau_j$  の分布についても同様の手続きで得た。

一様分布条件と等間隔条件は根本的に異なるサンプリングの手続きであるため、閾値についての異なる分布を生成する。一様分布条件は等間隔条件に比べて、より病的に（pathologically）整っていない閾値を生成する。例えば、 $m = 4$  のときに  $\{-\infty, 0.81, 0.82, 0.83, \infty\}$  といった閾値が可能なのは一様分布条件のみだからである。等間隔条件で生成可能な分布は全て一様分布条件からも生成可能である一方で、その逆は真でない。したがって、一様分布条件と等間隔条件を比較した場合に、一様分布条件の方がより整っていない閾値を生成するため、閾値が相関係数のバイアスへと及ぼす影響を検討することが可能となる。

いずれの閾値の発生条件においても  $[-1, 1]$  の範囲を用いたのは以下の理由からである。第 1 に、2.2 節で見たように分割表の一つのセルへの過集中は  $r$  の相当量の低下を招くが、例えば  $[-20, 20]$  のようなセルへの過集中や空白のセルを生じさせやすい大きすぎる範囲を設定した場合には、シミュレーションの設定は現実のデータ収集の状況とかけ離れたものとなり、あまり意味のないものになってしまうからである。さらに言えば、このような大きな範囲から発生させた閾値で離散化した場合には、分散が 0 となってしまいそもそも相関係数が計算できなくなるケースが多発してしまう。第 2 に、本研究のように閾値が相関係数の低下へと与える影響を検討する試みは今までに行われてきていないため、条件の設定においてはいくらか恣意的

な要素を含まざるを得ないことによる。すなわち、いずれかの値を設定しないことには検討そのものを始めることはできないということである。

### データ生成の手続き

各要因を組み合わせたデータ生成の条件は、13（母相関係数）× 21（カテゴリ数の組み合わせ）× 3（サンプルサイズ）× 20000（閾値各条件 10000 通り）で、全部で 1,638 万通りであった。それぞれの条件につき、まず 2 変量正規分布からデータを生成した後に変数をカテゴリ化し、そこから相関係数を計算することを 1000 回繰り返し、合計で 163 億 8000 万個の相関係数を算出した。

### ソフトウェア

シミュレーションは全て統計解析環境 R version 3.5.0(R Core Team, 2018)で行った。追加のパッケージとして 2 変量データの生成には mvtnorm version 1.0-8(Genz et al., 2018), 大規模データの扱いには ff version 2.2-14(Adler, Glaser, Nenadic, Oehlschlagel, & Zucchini, 2018) を用いた。

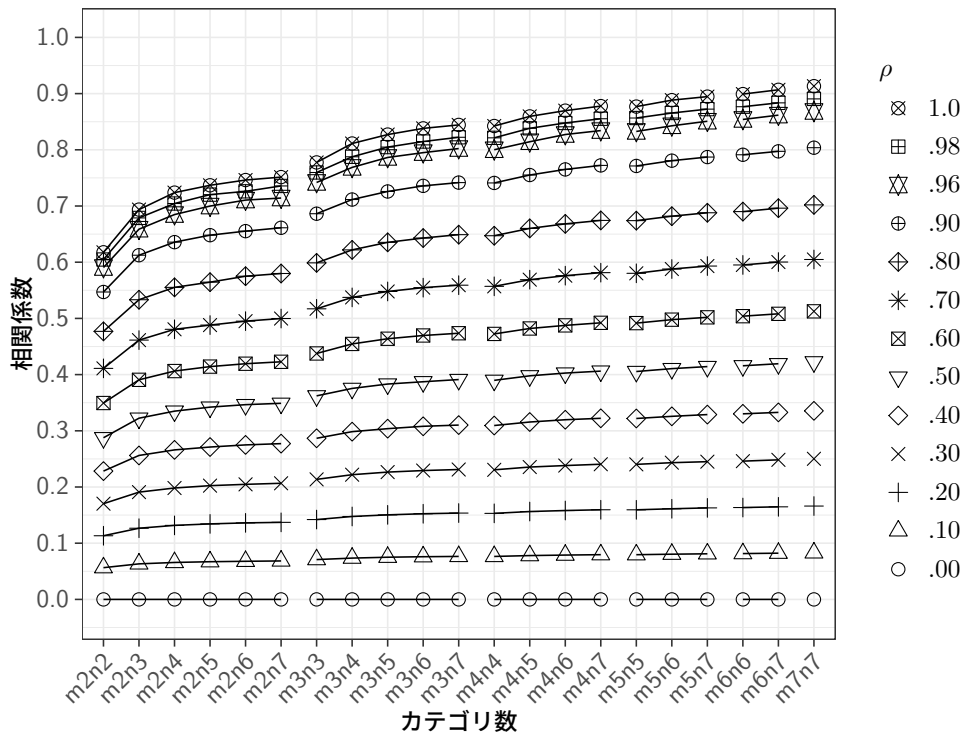
### 3.2.3 結果と考察

サンプルサイズが 1024 の場合の閾値の条件設定ごとの平均相関係数を Figure 3.1 に示した<sup>3</sup>。サンプルサイズ、カテゴリ数、閾値の条件に関わらず、 $r$  の値は  $\rho$  を一貫して過小推定していた。過小推定の程度はカテゴリ数が増えるとともに減少した。カテゴリ数が異なる組み合わせを用いた場合においてもこの傾向が見られた。予想されたことではあるが、バイアスの程度は 2 つの閾値の設定条件で大きく異なった。

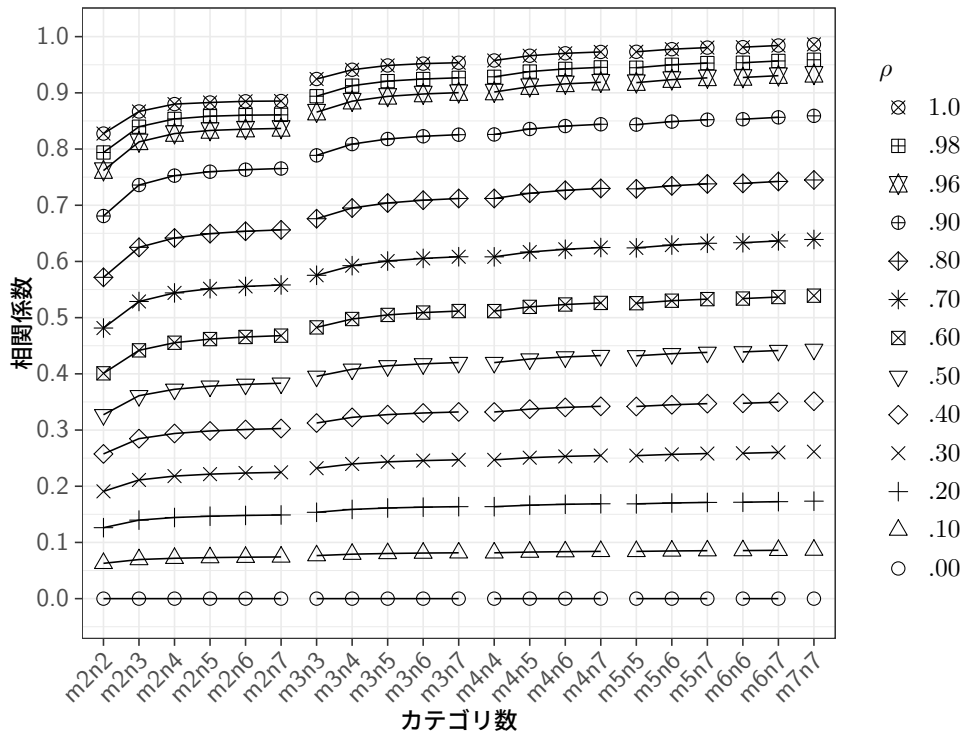
閾値設定の 2 つの条件を比べると、一様分布条件 (a) の方が等間隔条件 (b) よりもより深刻な  $r$  の低下を招いていた。例えば、 $m = 3, n = 4, \rho = .9$  でサンプルサイズ 1024 の場合では、 $r$  の平均は一様分布条件では、.712 である一方で、等間隔条件では .809 であった。区間を等分割するように閾値を設定した等間隔条件 (b) に比

---

<sup>3</sup> 平均相関係数については、サンプルサイズが 64, 512 の場合でも Figure 3.1 と同様の傾向であった。それぞれのサンプルサイズでの数表は付録 A に示した。



(a) 一様分布条件



(b) 等間隔条件

Figure 3.1. 各閾値設定条件における平均相関係数

べて、閾値が多様な値をとりうる一様分布条件 (a) ではより大きなバイアスが生じることが明らかであった。

また、シミュレーションの結果得られた値を 2 章の Figure 2.2 (b) で検討したカテゴリ化後の相関係数の期待値  $\rho_{XY}$  と比較すると、特別なケースである  $\rho = 1.0$  の場合を除いて等間隔条件との間には大きな差はなかった一方で、一様分布条件においては大きな隔たりが見られた。例えば、 $m = 4, n = 5, \rho = .8$  の場合には、Figure 2.2 (b) では  $\rho_{XY} = .726$  であるが、シミュレーションの結果の平均  $r$  は .660 であった。別の数値例では、 $m = 5, n = 7, \rho = .8$  では、Figure 2.2 (b) において  $\rho_{XY} = .741$  である一方で、シミュレーションの結果の平均  $r$  は .688 であった。

一様分布条件においてより大きなバイアスが生じた原因は、この条件設定がよく整った閾値と整っていない閾値の双方を含むからだと考えられる。例えば、閾値間の距離がとても近い場合には、分割表は空白（もしくはほぼ空白）のカテゴリを含む傾向にある。また、全ての閾値が範囲の上限（あるいは下限）に寄っている場合には分割表の値は左右非対称となる。これらのように、カテゴリ化の際にもとの分布の構造を破壊する変換が  $r$  の大きな低下を招いたと考えられる。他方、等間隔条件においてはそのような破壊的な変換は不可能である。したがって、2 つの閾値設定の間に見られた  $r$  の平均の差が、整っていない閾値を許容するか否かに起因すると考えることに無理はないだろう。

続いて、サンプルサイズの影響を検討するために、Figure 3.2 に一様分布条件における  $m = 6, n = 7$  の場合の  $r$  の分布を示した。ここからは、サンプルサイズを増やすと  $r$  の分散は小さくなるが、分布の中心的位置を動かさないことが明らかである。これが意味することは、サンプルサイズを増やしたところでカテゴリ化によって生じたバイアスを補正することはできず、バイアスがかかった値の推定精度を高めるのみだということである。また、これは同時に次の事実も示している。カテゴリ化は非線形変換であり変換によって新しい分布が得られるが、得られた新しい分布と変換前の分布から推定されるパラメータは一般に異なっているということである。

実際のデータ収集やデータ解析を念頭におくと、閾値の真の位置を知ることは非常に難しいと言える。しかし同時に、閾値の位置は回答者によって異なること、回答



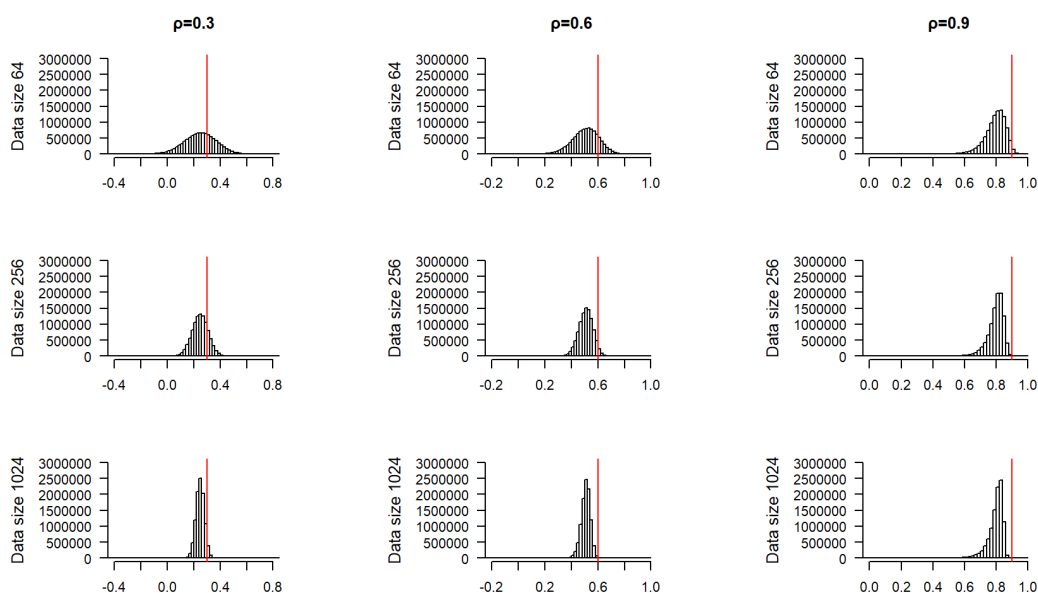


Figure 3.2.  $m = 6, n = 7$ における  $r$  の分布（垂直線は母相関係数）

の場面によって異なること、使われる項目によって異なることを仮定することは合理的だろう。したがって、シミュレーションの結果が示唆することは、多様な閾値の位置を考慮した場合に相関係数は従来考えられているよりも大きく低下するということである。

### 3.3 本章のまとめ

本章では、離散化による相関係数のバイアスの程度を検討した。先行研究ではカテゴリ化前の母相関係数が高くなるほどにバイアスが大きくなるという結論は一貫していたもの、カテゴリ数とバイアスの関係についてはカテゴリ数の増加が必ずしもバイアスの低減につながらないケースもあり一貫しないものであった。従来のシミュレーションではバイアスに影響を及ぼす重要な要因であるカテゴリ化の際の閾値が未検討であったことから、相関係数のバイアスの程度をより正確に評価するためにモンテカルロ・シミュレーションを実施した。シミュレーションの結果からは、変数の離散化の際の閾値が相関係数のバイアスへと与える影響は大きく、過去に離散化による相関係数のバイアスを評価したシミュレーションは、相関係数のバイア

スを過小評価している可能性があることを指摘した。Likert 型の項目から得られる順序カテゴリデータから計算される相関係数を用いる場合は、バイアスへの対処法を工夫しないと得られたデータを誤って解釈し、間違った結論を導く可能性があることが示唆された。

## 第 4 章

# カテゴリ化によって生じる相関係数のバイアスへの対処法の評価

本章では、カテゴリ化によって生じる相関係数のバイアスへの対処法の評価を行うことを目的とする<sup>1</sup>。バイアスへの対処法は大別すると、polychoric 相関係数を用いるものと Likert 型の項目への得点の与え方を工夫するものの 2 種類に大別できるので、4.1 節でそれぞれを概観する。続く 4.2 節では、それらの性能の評価を行うためのモンテカルロ・シミュレーションを行う（研究 2-a）。4.3 節では、シミュレーションの結果も踏まえて、実データを用いた場合にこれらの手法の違いがどのような差を生じさせるのかの検討も行う（研究 2-b）。

### 4.1 カテゴリ化によるバイアスへの対処法の先行研究

第 1 章で述べたように、カテゴリ化によって生じるバイアスの存在は古くから知られておりその対処法は 20 世紀前半から検討されてきた。Peters & Van Voorhis (1940) は相関係数の算出の際の分母の標準偏差に Sheppard の補正を用いることでバイアスを補正する式を提案したが、この方法ではバイアスを過修正してしまう場合があることが報告されており (Nishisato & Torii, 1971) , 簡易な対処法ではこの問題を上手く扱えないようである。バイアスへの対処法は大別すると polychoric 相

---

<sup>1</sup> 本章の内容は小野島・椎名 (2019) を加筆修正したものである。

関係数を用いるものと Likert 型の項目へと最適な得点を与えるものの 2 種類の方向性が考えられる<sup>2</sup>(椎名, 2016)。本節ではそれぞれを概観する。

#### 4.1.1 Polychoric 相関係数

##### Pearson による tetrachoric 相関とその応用

1 章でも述べたように、順序カテゴリデータから計算される相関係数のバイアスは、相関係数の提唱者である Karl Pearson 自身によって指摘されており、その対処法も早期から検討されていた。潜在変数が 2 変量正規分布に従うとの仮定のもとで、 $2 \times 2$  の分割表に対しては tetrachoric 相関係数が Pearson (1900) によって提案されている。Tetrachoric 相関係数では、周辺比率を用いて 2 変量正規分布を 4 分割する 2 直線の交点  $(h, k)$  を求め、それらの累積確率を用いて元の 2 変量正規分布の相関係数を推定する。後にこの計算を簡便にするための数表が Everitt (1910) によって提供されている。

Ritchie-Scott (1918) は tetrachoric 相関係数をベースにして、polychoric 相関係数を提案している。Ritchie-Scott の polychoric 相関係数では、 $r \times c$  の分割表から 2 変量正規分布を分割する  $r - 1$  と  $c - 1$  本の直線を考える。それぞれの交点からは交点に対応する 4 象限 (quadrant) を考えることができるので、交点の数だけ tetrachoric 相関係数を考えることができる。平面の交点の数、すなわち  $r - 1 \times c - 1$  だけ存在する 4 象限からそれぞれ tetrachoric 相関係数を算出して、それらの確率誤差が最小となるように重み付きの平均を求めることで、2 変量正規分布の相関の近似とする手法である。

Ritchie-Scott (1918) の手法は当時では計算に大変労力がかかるものであったため、Pearson & Pearson (1922) は計算の労力の少ない手法を提案し、新手法でも遜色ない正確性で相関係数を推定できると主張している。しかしながらこの手法は級数を用いる複雑なものでもあったことから心理学全体としては広くは使われず、

---

<sup>2</sup> Pearson (1913) はカテゴリ化によって希薄化した相関係数の補正式を提案したが、現在までにこの補正式の適用例は極めて少なく、また、その後の理論的な発展もあまり見られないため本節では検討しない。なお、最近 Shiina, Ueda, & Kubo (2018) がこのアイデアを発展させて EM アルゴリズムによる polychoric 相関係数を提案している。

polychoric 相関係数が実用的に用いられるようになるのは Olsson (1979a) の提案以降であった。

### 最尤法を用いた Olsson の方法

Olsson (1979a) による polychoric 相関係数は、カテゴリカルデータの背後に 2 変量正規分布を仮定して、連続変量がカテゴリへと区切られる閾値および 2 変量正規分布の母相関係数について最尤推定を行うものである。Olsson は閾値と母相関係数を同時に推定する 1 ステップ法と、閾値は周辺累積比率を用いて推定した後に母相関係数のみを最尤推定する 2 ステップ法を紹介しているが、実用的には後者が使われることが多いようである。

ここで、2 ステップ法の手続きの概略を紹介する。例として、 $\rho = .8$  の 2 変量標準正規分布から発生させた  $n = 1000$  のデータを適当な閾値で離散化した 2 変量  $X$  と  $Y$  が Table 4.1 のように得られているものとする。

Table 4.1  
分割表の例

		Y				行合計
		1	2	3	4	
X	1	34	30	0	0	64
	2	36	306	95	0	437
	3	0	84	310	30	424
	4	0	1	29	45	75
列合計		70	421	434	75	1000

Olsson の 2 ステップ法では、最初のステップとして周辺累積比率から閾値が推定される。すなわち、 $P_i$  と  $P_j$  をそれぞれ  $X$  と  $Y$  の周辺累積比率として累積分布関数の逆関数  $\theta_i = \Phi^{-1}(P_i)$ ,  $\tau_j = \Phi^{-1}(P_j)$  が用いられる。Table 4.1 の累積比率を求めると、 $X$  について  $[0.064, 0.501, 0.925, 1.000]$  であるため閾値は  $\theta_1 = -1.522, \theta_2 = -0.003, \theta_3 = 1.440, \theta_4 = \infty$  となる。 $Y$  についても同様に、 $\tau_1 = -1.476, \tau_2 = -0.023, \tau_3 = 1.440, \tau_4 = \infty$  である。 $\theta_0$  と  $\tau_0$  については  $-\infty$  とする。

続くステップ 2 では、推定した閾値と分割表を用いて母相関係数の最尤推定が行われる。 $n_{ij}$  を  $X = i$  かつ  $Y = j$  である観測度数、 $\pi_{ij}$  を 2 変量標準正規分布の元で  $(i, j)$  のセルに落ちる確率とすると、尤度関数は

$$l = \ln K + \sum_{i=1}^m \sum_{j=1}^n n_{ij} \ln \pi_{ij} \quad (4.1)$$

として表される。このとき  $\pi_{ij} = \Phi_2(\theta_i, \tau_j) - \Phi_2(\theta_{i-1}, \tau_j) - \Phi_2(\theta_i, \tau_{j-1}) + \Phi_2(\theta_{i-1}, \tau_{j-1})$  である。(4.1) を  $\rho$  で偏微分すると解くべき式は、

$$\frac{\partial l}{\partial \rho} = \sum_{i=1}^m \sum_{j=1}^n \frac{n_{ij}}{\pi_{ij}} [\phi_2(\theta_i, \tau_j) - \phi_2(\theta_{i-1}, \tau_j) - \phi_2(\theta_i, \tau_{j-1}) + \phi_2(\theta_{i-1}, \tau_{j-1})] = 0 \quad (4.2)$$

となる。Table 4.1 のデータに対して、尤度を最大化する値を Newton 法を用いて実際に求めると  $\hat{\rho} = .838$  となる。

#### 4.1.2 Likert 型の項目に整数値以外の値を割り当てる方法

O'Brien (1981, 1983, 1985) は、潜在的な連続変数がカテゴリ化された上で分析される際に生じる誤差を「グループ化誤差 (grouping error)」と呼び、グループ化誤差は「カテゴリ化誤差 (categorization error)」と「変換誤差 (transformation error)」の 2 つからなるものとしてそれぞれを検討している。前者は、1 つのカテゴリに属する連続変数が全て同じ値にされてしまうことによって生じる誤差で、後者はカテゴリ化後に割り当てられる値が連続変数のカテゴリにおける平均と異なることによって生じる誤差であるとし、後者の誤差はカテゴリ化の際のスコアリングの仕方によって減少することを報告している。

Likert 型の項目を用いたほとんどの研究は等間隔の [1, 2, 3, 4] といった整数値を与える方法を用いているが、順序カテゴリデータの各カテゴリ値として等間隔でない数値をデータの構造から計算して割り当てる方法が今までに考案されている。ここでは、Likert (1932) による sigma 法、Bross (1958) による ridity 法、西里 (1982) による双対尺度法を扱う。

## Sigma 法

あまり知られていないことではあるが現在のように Likert 型の項目に整数値を与える方法は簡便法として提案されたものである。Likert (1932) は簡便法と同時に sigma 法を提案し、両者の相関が非常に高いことから簡便法の利用の正当性を主張していたのである。Sigma 法ではカテゴリ得点は以下のように求められる。まず、カテゴリの累積比率の値と、累積正規分布の逆関数を用いて、カテゴリの境界値（閾値）を求める。その後、求めた境界値間で定義される切断正規分布の期待値をカテゴリ得点とする。なお、切断正規分布の期待値  $\bar{X}$  は、標準正規分布の確率密度関数を  $\phi(x)$ 、累積分布関数を  $\Phi(x)$ 、分布の下限と上限をそれぞれ  $a, b$  としたとき、

$$\bar{X} = \frac{\phi(a) - \phi(b)}{\Phi(b) - \Phi(a)} \quad (4.3)$$

として求めることができる (Guilford, 1954)。例として Table 4.1 でのカテゴリ化された変数  $X$  について sigma 法を用いてカテゴリ得点を求めると、まず累積比率が  $[0.064, 0.501, 0.925, 1.000]$  であり、その値をもとに累積分布の逆関数から境界値を計算すると、 $[-1.522, 0.003, 1.440, \infty]$  が求まる。1 番目のカテゴリの下側境界値は  $-\infty$  とし、(4.3) に従い切断正規分布からの期待値を求めると、カテゴリ得点はそれぞれ  $[-1.957, -0.626, 0.607, 1.887]$  となる。この値を  $[1, 2, 3, 4]$  の代わりに使うことになる。

## Ridit 法

Ridit 法とは、もとは Bross (1958) により考案されたもので、現在では経済学や医療のデータ解析の場面で使われることが多いようだが、心理や教育データへ適用している例は少ない。Bross によれば、ridit という名前は logit や probit などのアナロジーであり、他と同じく変換の一種を表しているためにつけられた名前である。各カテゴリに対応する ridit の値は次のようにして求められる。カテゴリ数が  $k$  であるときに、 $i$  番目のカテゴリの比率を  $p_i (i = 1, \dots, k)$  とすると、 $i$  番目のカテゴリに

割り当てられる ridit の値は

$$R_i = \frac{1}{2} \left( \sum_{j < i} p_j - \sum_{j > i} p_j \right) + 0.5 \quad (4.4)$$

で求めることができる (Brockett & Levine, 1977)。例えば Table 4.1 の  $X$  を再度とりあげると、 $X$  の 1 番目のカテゴリの比率は、0.064 であるため、式より、 $R_1 = (1/2) \times (0 - 0.936) + 0.5 = 0.032$  と求まる。同様に計算を行うと、 $X$  のカテゴリのスコアはそれぞれ  $[0.032, 0.283, 0.713, 0.963]$  となり、この値を  $[1, 2, 3, 4]$  の代わりに使うことになる。このように ridit は 0 から 1 の値をとる。

## 双対尺度法

ここまでの方法では分割表の周辺比率をもとにしてスコアリングを行っていたが、2 変量間の関係性に注目してスコアリングを行う手法も存在している。西里 (1982) の双対尺度法では、データの構造から最適な重み (得点) を求めるアルゴリズムが考案されている<sup>3</sup>。双対尺度法のアルゴリズムの一つである交互平均法を用いて、分割表の重みを計算する手順は以下の通りである (手順の詳細は西里 (1982) を参照されたい)。

- ステップ 1: 分割表の列に任意の重みを与え、行の平均ベクトル  $\mathbf{x}_1^*$  を求める。
- ステップ 2:  $\mathbf{x}_1^*$  で重みづけた得点の総和が 0 になるように変換し  $\mathbf{x}_{10}$  を得る。
- ステップ 3:  $\mathbf{x}_{10}$  の要素を最大の絶対値で割って基準化したものを  $\mathbf{x}_1$  とする。
- ステップ 4:  $\mathbf{x}_1$  を行の重みとして、列の平均ベクトル  $\mathbf{y}_1^*$  を求める。
- ステップ 5:  $\mathbf{y}_1^*$  で重みづけた得点の総和が 0 になるように変換し  $\mathbf{y}_{10}$  を得る。
- ステップ 6:  $\mathbf{y}_{10}$  の要素を最大の絶対値で割って基準化したものを  $\mathbf{y}_1$  とする。

この後は、ステップ 6 で得られた  $\mathbf{y}_1$  を再度ステップ 1 で重みとして代入し  $\mathbf{x}_2^*$  を求め、次いで原点を調整した  $\mathbf{x}_{20}^*$ 、それを基準化した  $\mathbf{x}_2$  を求めるといったように、この過程を繰り返す。行列の重みが安定するまで  $i$  回繰り返したものをそれぞれ  $\mathbf{x}_i, \mathbf{y}_i$

---

<sup>3</sup> なお、双対尺度法は、他にも「交互平均法」「符号解析法」「最適尺度法」などの名称でも呼ばれている。



として、最適な重み  $\mathbf{x}^*, \mathbf{y}^*$  は、

$$\mathbf{x}^* = \left[ \frac{f_t}{{}^t\mathbf{x}_i D \mathbf{x}_i} \right]^{\frac{1}{2}} \mathbf{x}_i, \quad \mathbf{y}^* = \left[ \frac{f_t}{{}^t\mathbf{y}_i D_n \mathbf{y}_i} \right]^{\frac{1}{2}} \mathbf{y}_i \quad (4.5)$$

で求めることができる。ここで、 $f_t$  は総度数、 $D$  は行の周辺度数を対角項にもつ対角行列、 $D_n$  は列の度数を対角項にもつ対角行列である。

例として Table 4.1 の分割表に双対尺度法を適用して最適な重みを求めると、 $\mathbf{x}^* = {}^t(-1.816, -0.750, 0.661, 2.185)$ 、 $\mathbf{y}^* = {}^t(-1.776, -0.753, 0.636, 2.206)$  となる。これらの値を [1, 2, 3, 4] の代わりに使うことになる。なお、ここで与えられた重み（得点）から計算される相関係数は最大値を与えることが証明されている。すなわち、2 変量に与える重みの組み合わせには様々な値を考えることができるが、それらの重みの組み合わせの内、相関係数を最大化する基準で選ばれた重みの組み合わせは、双対尺度法の手順で求めた重みの組み合わせと一致するということである (岩坪, 1987; Kendall & Stuart, 1979)。

## 4.2 カテゴリ値に整数値以外を割り振った場合および polychoric 相関係数の比較 (研究 2-a)

### 4.2.1 問題と目的

前節で、Likert 型の項目から得られる順序カテゴリデータから相関係数を推定する際の手法として、polychoric 相関係数を用いるものと、各カテゴリに整数値を与える代わりに異なる値を与える方法を紹介した。Polychoric 相関係数については、Olsson (1979a) が簡単な性能の評価を行っているが、Likert 型の回答に整数値以外の得点を与えた場合に離散化による相関係数のバイアスがどのように影響を受けるのかの検討はいまだされていない。そこで本節では相関係数のバイアスへの対処法として、polychoric 相関係数、および、カテゴリに整数値とは異なる値を割り当てたのちに計算される相関係数の比較を目的とした。

## 4.2.2 方法

シミュレーションでは、(a) 母相関係数、(b) カテゴリ数、(c) 離散化の際の閾値、(d) サンプルサイズの4つの要因を操作して、順序カテゴリデータを生成した。各条件で生成されたデータに対して(1) Likertの簡便法、(2) sigma法、(3) ridit法、(4) 双対尺度法で数値を与えたものから相関係数を計算し、また(5) polychoric相関係数も算出して比較した。

### 母相関係数

研究1と同様に、母相関係数については、低い相関から高い相関まで網羅的に検討できるように、水準を0, .1, .2, .3, .4, .5, .6, .7, .8, .9, .96, .98, 1.0として設定した。

### カテゴリ数

研究1と同様に、応用研究において使われることが多い2カテゴリから7カテゴリまでの全ての組み合わせの水準を設定した。

### 閾値の設定

応用場面で連続変数は多様な閾値によって離散化されていると考えられるため、本来であれば研究1のシミュレーションで行なったように多様な閾値を生成して、それらを用いて連続変数を離散化することが望ましい。しかしながら、本シミュレーションにおいては相関係数の計算において反復計算が必要な手続きが含まれるため、データ生成の条件の組み合わせ数が増えるとシミュレーションに膨大な時間がかかり実施が現実的ではなくなる。そこで2.2節での検討も踏まえて、カテゴリ化後の変数の歪みの程度と、2変量の歪む方向を操作して条件を設定した。具体的には、左右対称、同一方向の歪み(弱)、同一方向の歪み(強)、逆方向の歪み(弱)、逆方向の歪み(強)の5つの水準を設定した。

左右対称条件では、離散化後のカテゴリ変数の比率が等比率になるように閾値を設定した。同一方向の歪み(中)条件では、例えば3カテゴリの場合にカテゴリ変数

の比率が  $\{0.5, 0.3, 0.2\}$  と単調に減少していくように 2 変量の閾値を設定した<sup>4</sup>。同一方向の歪み（強）条件では、 $\{0.7, 0.2, 0.1\}$  とより程度の大きい歪みを設定した。逆方向の歪み（中）条件では、 $X$  に対しては同一方向の歪み（中）の条件で用いられた閾値と同じ値を、 $Y$  ではその正負を反転させたものを用いた。逆方向の歪み（強）も同様の手続きを用いて作成した。同一方向の歪みと逆方向の歪みの例を Figure 4.1 に示し、各水準において実際に用いた閾値を Table 4.2 に示した。

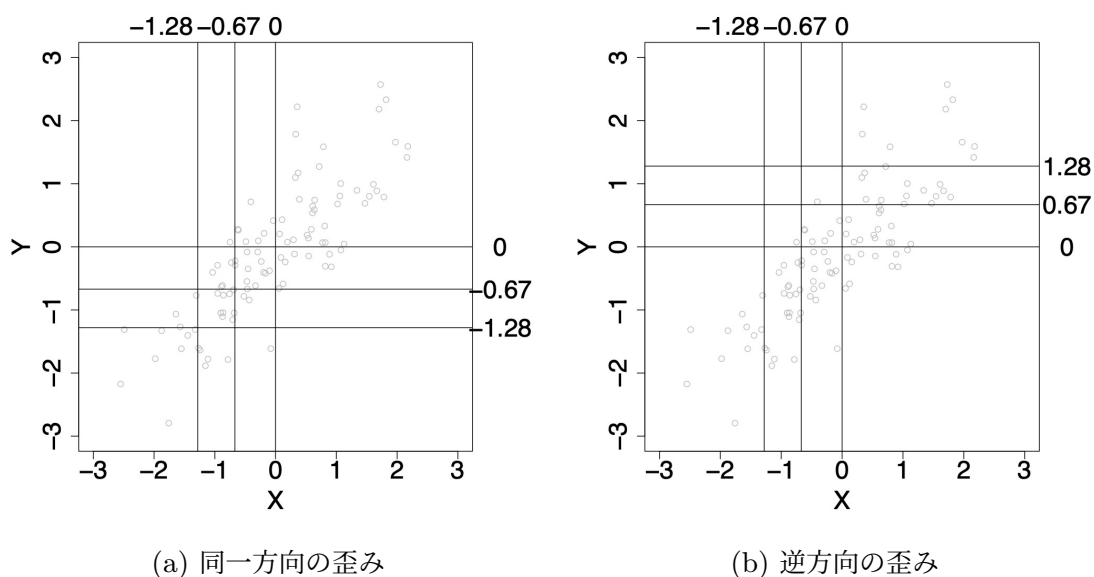


Figure 4.1. 用いた閾値の例（カテゴリ数が 4 で強い歪み）

### サンプルサイズ

小さいサンプルサイズから大きなサンプルサイズまで検討できるように 50, 100, 200, 500, 1000 の 5 水準を設定した。

### データ生成の手続きと相関係数の計算

以上のデータ生成の要因を組み合わせたセル数は、13(母相関係数) × 21(カテゴリ数) × 5(閾値) × 5(サンプルサイズ) = 6825 であった。各条件に

<sup>4</sup> カテゴリ変数が与えられたような比率になるような閾値は分布関数の逆関数から求めることができる。例えば比率が  $\{0.5, 0.3, 0.2\}$  であれば閾値は  $\{0, 0.842\}$  である。

Table 4.2  
研究 2 で離散化の際に用いた閾値

カテゴリ数	歪みの程度	閾値
2 カテゴリ	左右対称	0
	弱い歪み	-0.25
	強い歪み	-0.84
3 カテゴリ	左右対称	-0.43, 0.43
	弱い歪み	-0.67, 0.13
	強い歪み	-1.04, -0.25
4 カテゴリ	左右対称	-0.67, 0, 0.67
	弱い歪み	-1.28, -0.52, 0.25
	強い歪み	-1.28, -0.67, 0
5 カテゴリ	左右対称	-0.84, -0.25, 0.25, 0.84
	弱い歪み	-1.64, -0.84, -0.25, 0.39
	強い歪み	-1.64, -1.03, -0.52, 0
6 カテゴリ	左右対称	-0.97, -0.43, 0, 0.43, 0.96
	弱い歪み	-1.88, -1.28, -0.67, -0.12, 0.52
	強い歪み	-2.05, -1.41, -0.84, -0.39, 0.13
7 カテゴリ	左右対称	-1.07, -0.57, -0.18, 0.18, 0.57, 1.06
	弱い歪み	-2.05, -1.64, -1.03, -0.61, -0.13, 0.52
	強い歪み	-2.33, -1.88, -1.28, -0.84, -0.39, 0.13

つき、それぞれ 500 のデータセットを 2 変量正規分布から生成した。まず生成された連続データに対して、(2.3) のように閾値を適用して変数をカテゴリ化した。その後、離散化されたデータに対して (1) Likert の簡便法, (2) sigma 法, (3) ridit 法, (4) 双対尺度法のそれぞれの手法でカテゴリ値を割り当てた後に、Pearson の相関係数を計算し記録した。また、離散化されたデータから (5) polychoric 相関係数を計算した。Polychoric 相関係数の計算の際には、分割表の周辺比率から閾値を求めた後に母相関係数のみを最尤推定する 2 ステップ法を用いた。

## ソフトウェア

シミュレーションは全て統計解析環境 R version 3.6.3(R Core Team, 2020)で行った。追加のパッケージとして、2変量データの生成には `mvtnorm` version 1.1-0(Genz et al., 2020), polychoric 相関係数の算出には `polycor` version 0.7-10(Fox, 2019) に所収の関数を使用した。Sigma 法, ridit 法, 双対尺度法についてはそれぞれ自作の関数を用いた。

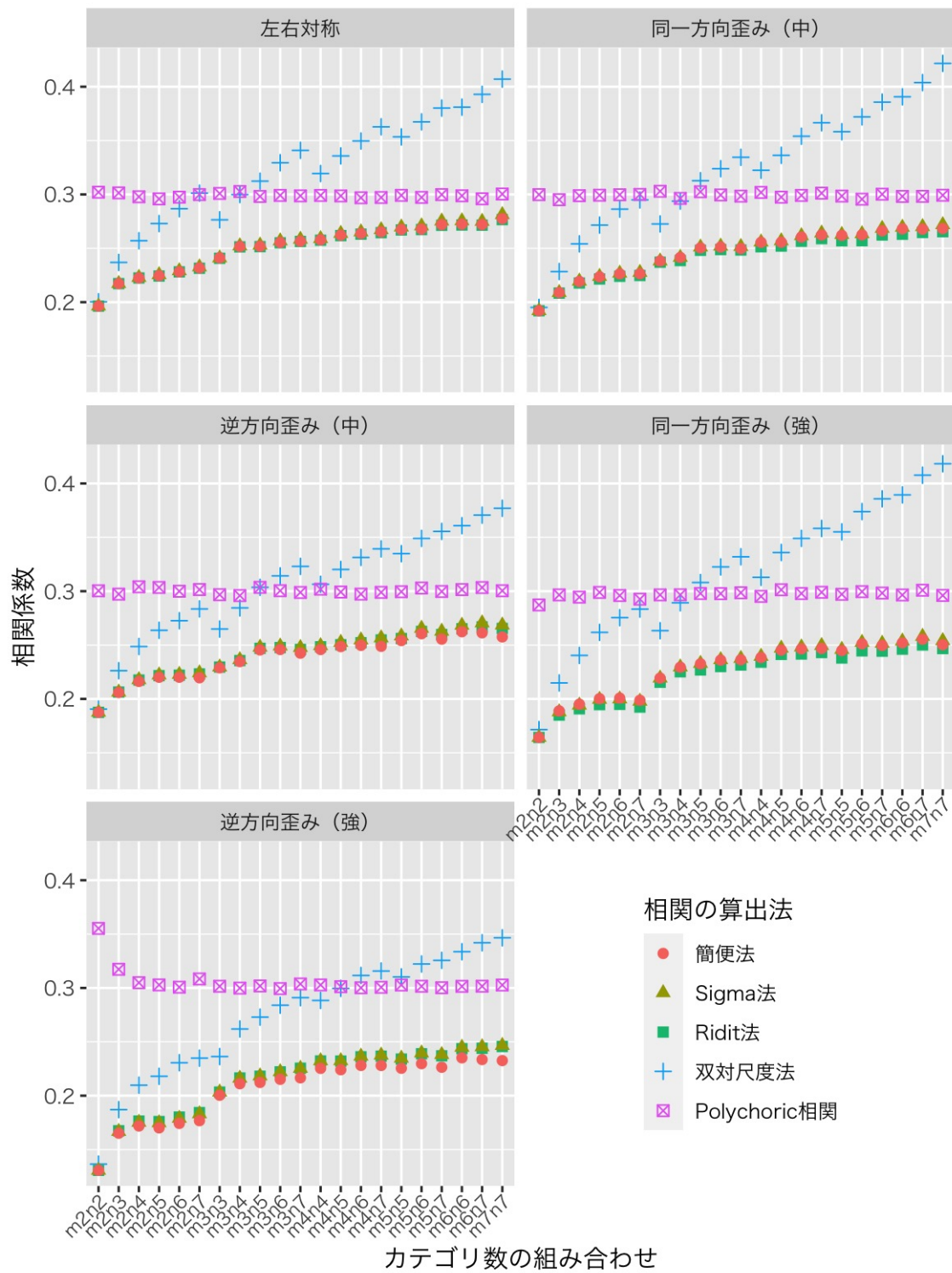
### 4.2.3 結果と考察

双対尺度法について、データ行列によっては解が存在せず交互平均法で重みを計算できない場合があるが、そうした場合には計算から除外した。交互平均法で計算が不能になるケースが生じたのは  $m = n = 2$  および  $m = 2, n = 3$  の場合のみで、これらの全ての条件下で発生率は 1% 以下であった。

まず相関係数算出の手法ごとのバイアスを比較するために、全てのサンプルサイズの結果を合算したデータに対して、母相関係数、閾値の条件、カテゴリ数の組み合わせ、相関係数の算出法毎に相関係数の平均を求めた。結果の全てを示すことは冗長であるため、ここでは  $\rho = .3, \rho = .6, \rho = .9$  の場合のみを Figure 4.2 に示し、その他の母相関係数の結果については付録 B に示す。

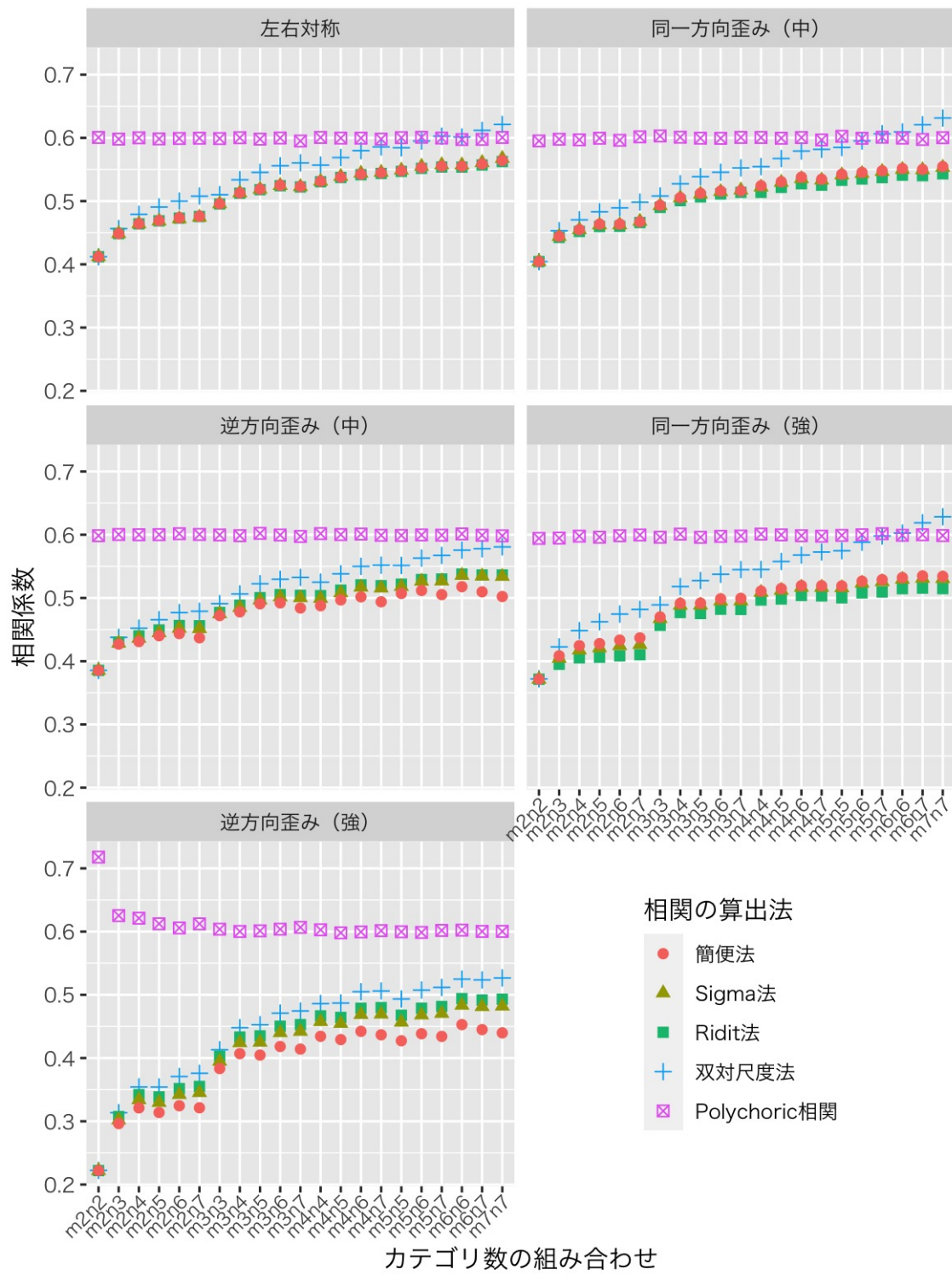
相関係数の算出法を比較をすると、シミュレーションの多様な条件において polychoric 相関係数をもっとも正確に母相関係数を推定することができていた。Polychoric 相関係数による推定が上手くいかない場合は、 $m = n = 2$  や  $m = 2, n = 3$  などのように用いられるカテゴリ数が小さく、かつカテゴリ化後の 2 変量が逆方向に強く歪んでいる場合であった。このとき polychoric 相関係数には過大推定のバイアスが生じた。また、 $\rho = .96$  や  $\rho = .98$  のように元の相関が非常に高く、かつカテゴリ数が大きい時には過小推定のバイアスが生じた (Figure B.8-9)。

Likert の簡便法, sigma 法, ridit 法のいずれも、絶対値が低下するバイアスが生じており、前章でのシミュレーションと同様にカテゴリ数が増加するとともにバイアスの程度は小さくなっていった。この 3 つの手法については、閾値条件が左右対



(a)  $\rho = .3$  の場合

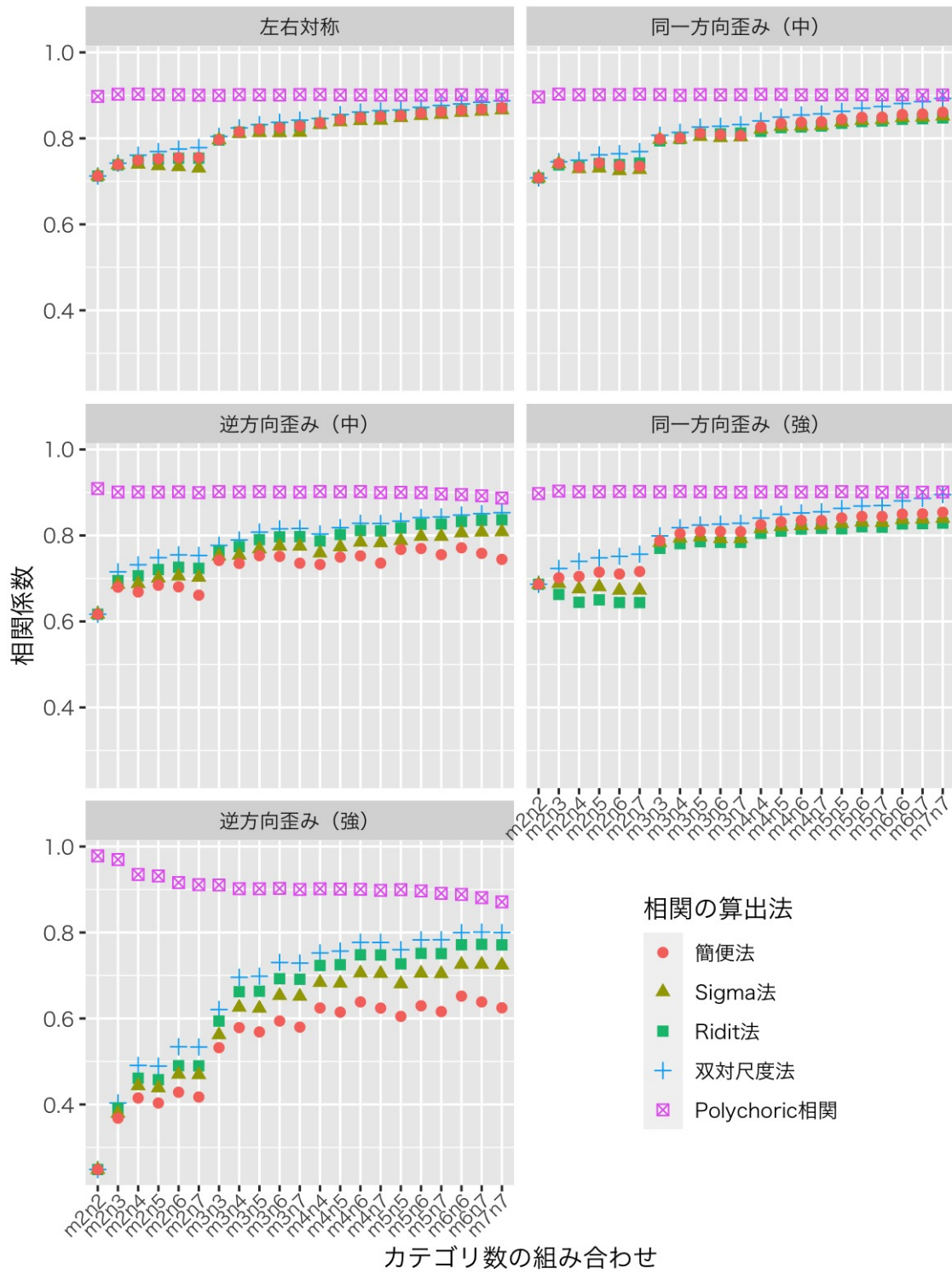
Figure 4.2. 平均相関係数



カテゴリ数の組み合わせ

(b)  $\rho = .6$  の場合

Figure 4.2. 平均相関係数 (続き)



カテゴリ数の組み合わせ

(c)  $\rho = .9$  の場合

Figure 4.2. 平均相関係数 (続き)



称の場合や、カテゴリ化後の分布の歪みが大きくない場合には、ほぼ同様の挙動を示していたが、2変量が逆方向に強く歪んでいる場合には Likert の簡便法において特に大きなバイアスが生じる傾向があった。

双対尺度法については、カテゴリ数の増加とともに相関係数が高くなっていく傾向が見られた。 $\rho = .3$  程度までの低い母相関の場合には、カテゴリ数が小さい場合には過小推定、カテゴリ数が大きくなった場合には過大推定をする傾向にあった。 $\rho = .7$  よりも高くなると、カテゴリ数が多い場合でも過大推定することはなくなり、カテゴリ数が増加するにつれて真の値に近づいていく傾向が見られた。Likert の簡便法、sigma 法、ridit 法に比べると高い値を算出する傾向にあった。

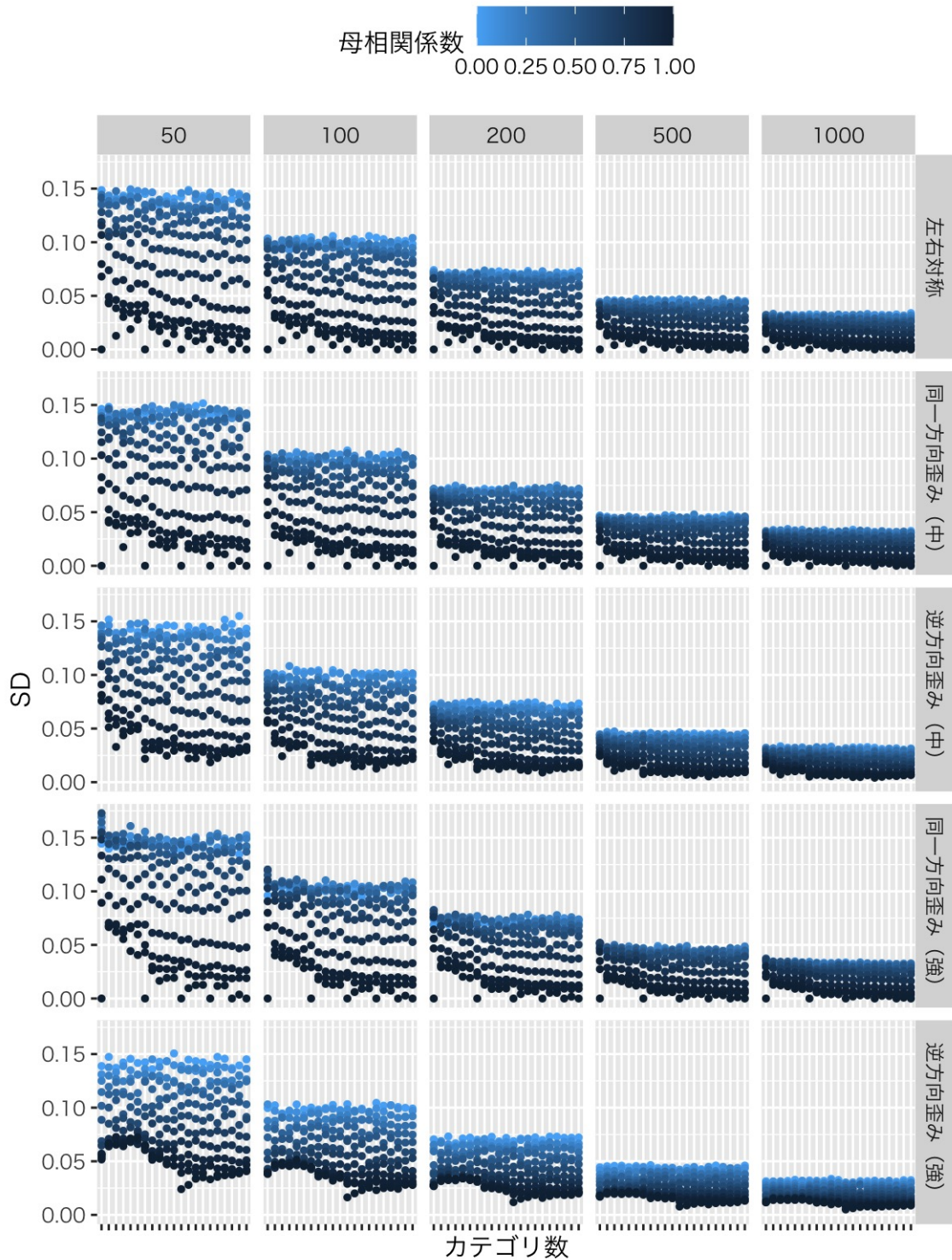
続いて、サンプルサイズの影響を検討するために、各条件ごとに相関係数の SD を算出したものを相関係数の算出法別に Figure 4.3 に示した<sup>5</sup>。ここからは全ての相関の算出法に共通する傾向として以下の3点が指摘できる：(1) サンプルサイズが大きいかほど相関係数のばらつきは小さくなる、(2) 母相関係数が高いほど相関係数のばらつきは小さくなる、(3) 離散化後の分布が歪むほど、相関のばらつきは大きくなり、2変量の歪みの方向が逆方向な場合にばらつきはさらに大きくなる。

Likert の簡便法、sigma 法、ridit 法については、母相関係数が高い場合にはカテゴリ数の増加とともにばらつきは小さくなるが、母相関係数が低い場合にこの傾向は見られなかった。Polychoric 相関係数については、カテゴリ数の増加とともにばらつきが小さくなる傾向が見られた。双対尺度法については、離散化後の2変量が同一方向に歪んでいる場合には、カテゴリ数が増えるとともにばらつきは一旦小さくなるものの、さらにカテゴリ数が増えた場合にばらつきが大きくなる傾向が見られた。左右対称、逆方向の歪みの場合にはこの傾向は見られなかった。

分布の中心的位置については、Likert の簡便法、sigma 法、ridit 法、polychoric 相関係数ではサンプルサイズが増えても変わらなかった。その一方で、双対尺度法ではサンプルサイズが大きくなると平均相関係数が小さくなる傾向が見られた。離散化による絶対値減少のバイアスとサンプルサイズの増加に伴う影響が組み合わさっ

---

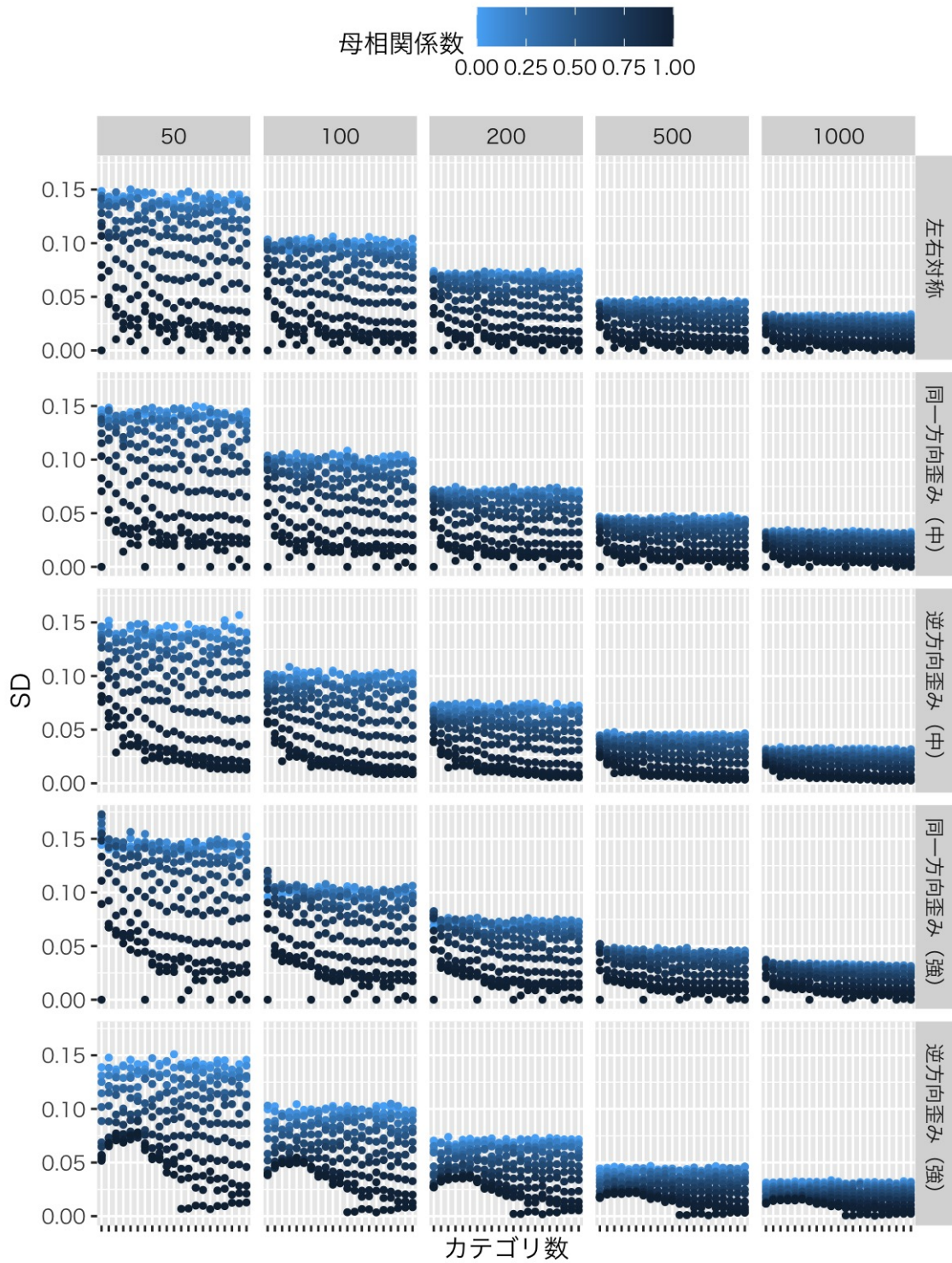
<sup>5</sup> なお、Figure 4.3 では、(e)polychoric 相関係数のみ Y 軸のスケールが違うことに注意が必要である。これは、カテゴリ数が少なく分布の歪みが強い場合には polychoric 相関係数の推定は不安定になることから生じている。



(a) Likert の簡便法

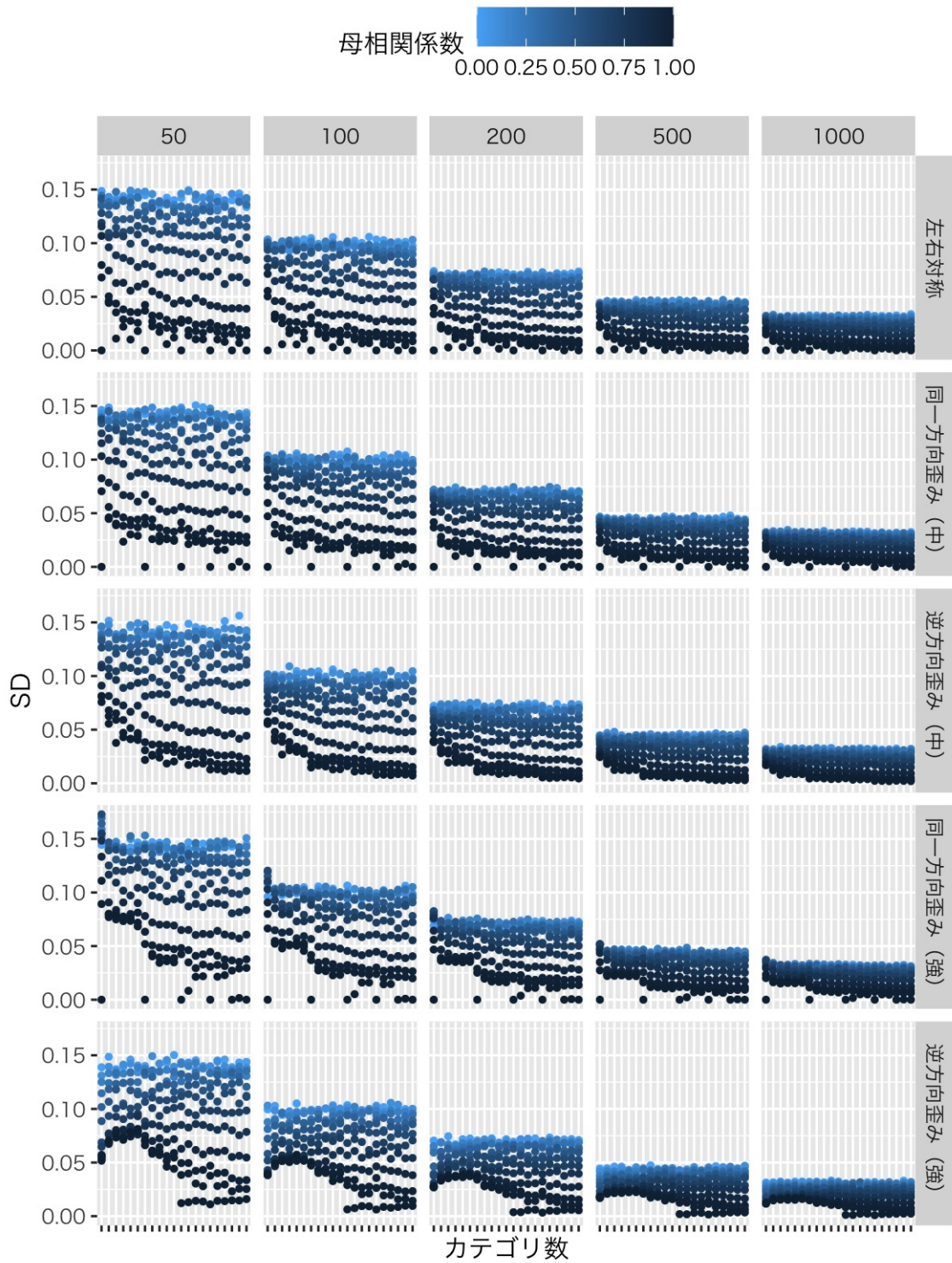
Figure 4.3. 条件毎の相関係数の SD

注) カテゴリ数の並びは Figure 4.2 と同じで左から  $m_{2n2}$ ,  $m_{2n3}$ , ...,  $m_{3n3}$ , ...,  $m_{7n7}$  である。



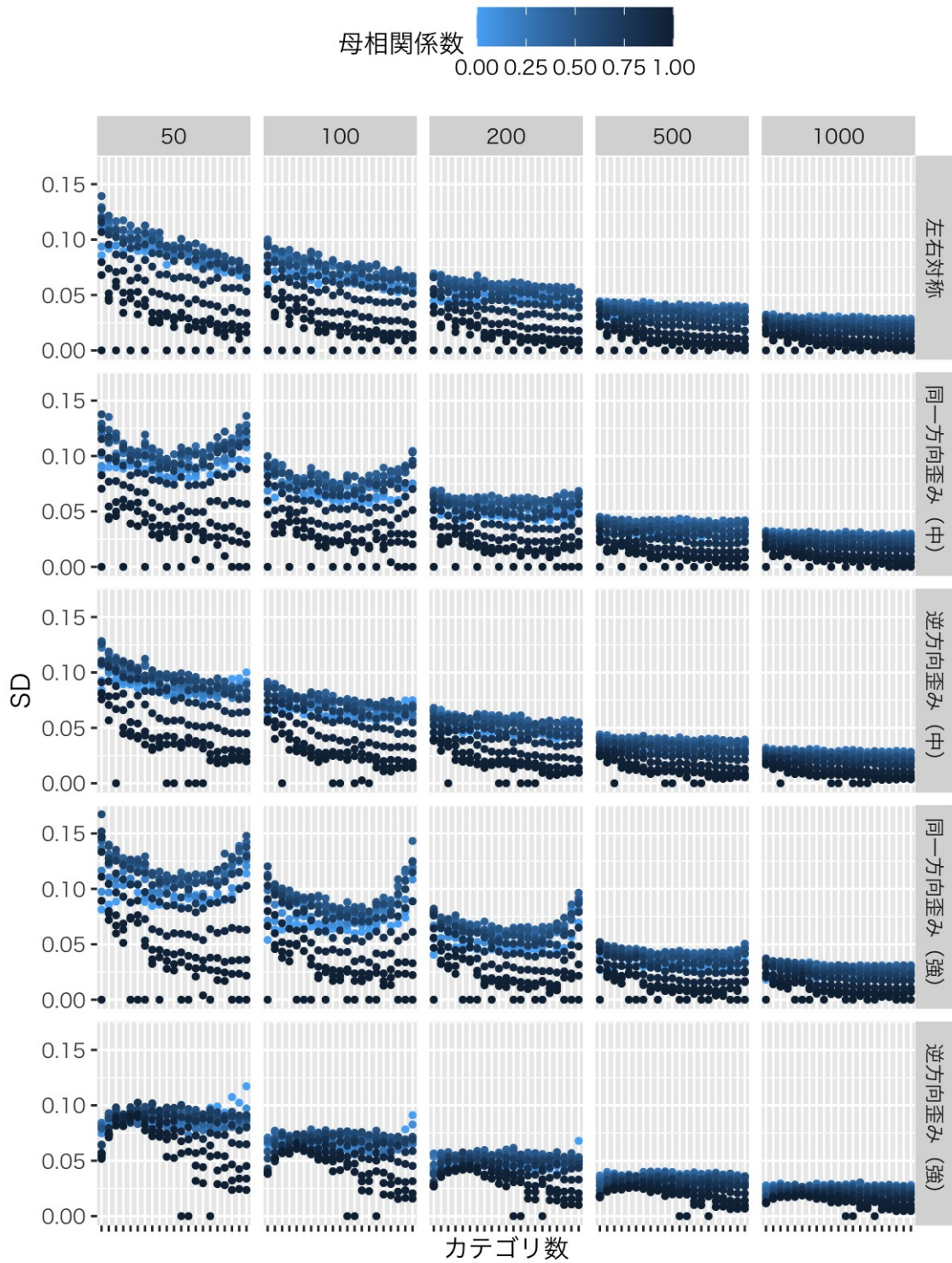
(b) Sigma 法

Figure 4.3. 条件毎の相関係数の SD (続き)



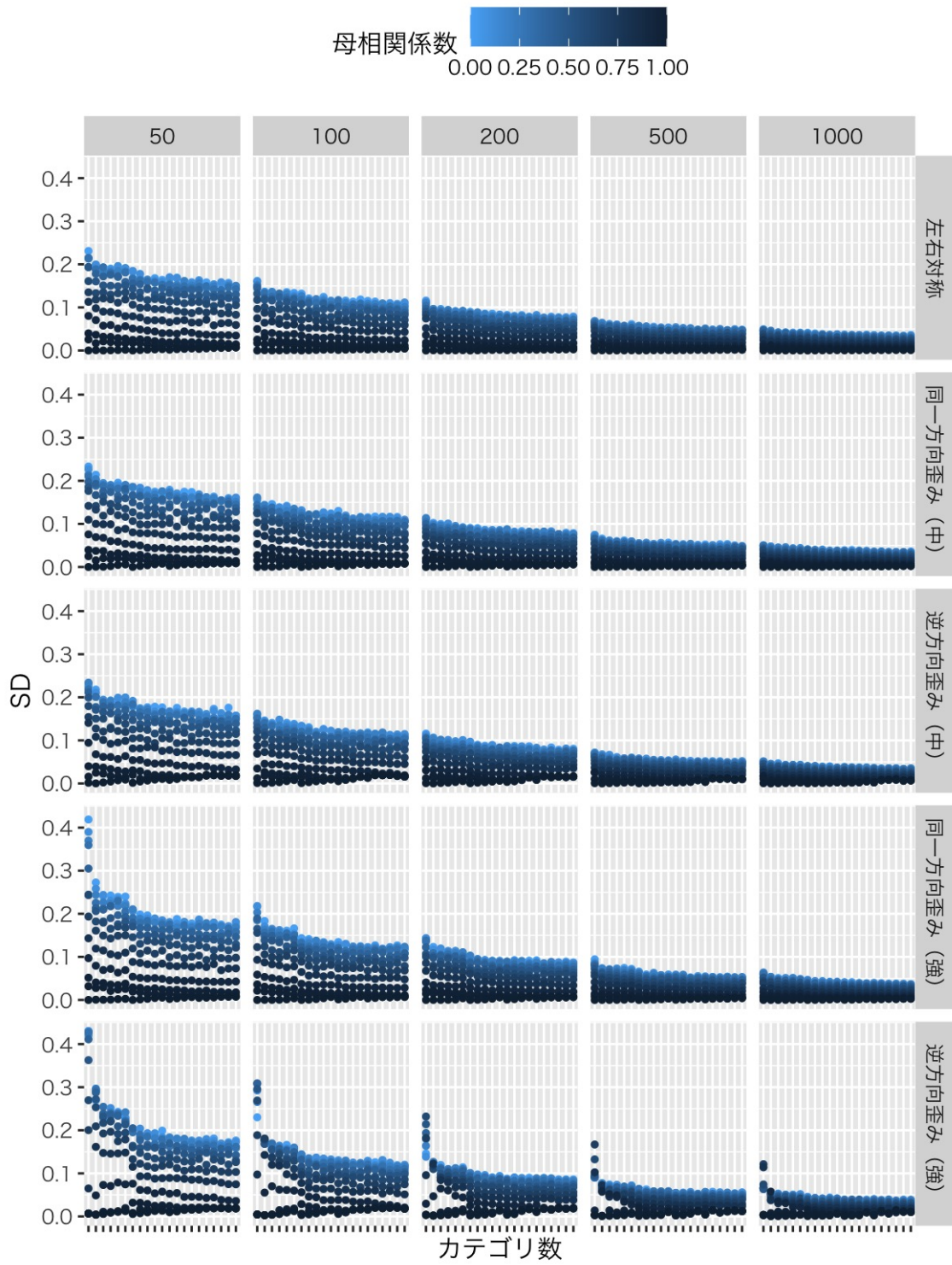
(c) Ridit 法

Figure 4.3. 条件毎の相関係数の SD (続き)



(d) 双対尺度法

Figure 4.3. 条件毎の相関係数の SD (続き)



(e) Polychoric 相関

Figure 4.3. 条件毎の相関係数の SD (続き)

て、母相関係数が低い時にはサンプルサイズが増えるとともに真値から遠ざかる傾向に、母相関係数が高い時には真値に近づく傾向がみられた。

以上の結果から、離散化によって生じたバイアスへの対処法を検討するときは、polychoric 相関係数が第一の選択肢として適していると結論できる。シミュレーションで扱った相関係数の算出法の中では、もっとも安定して母相関係数を推定することができており、応用研究のさまざまな場面で安定した活用が可能であろう。特に、カテゴリ化後の 2 変量の分布が歪んでいる場合には Likert の簡便法ではたとえカテゴリ数が多い場合であってもそれなりに大きなバイアスが生じるため polychoric 相関係数の利用が望ましいだろう。

Polychoric 相関係数を用いてもバイアスが生じるため利用に注意が必要な場合として、カテゴリ化後の 2 変量の分布が逆方向に強く歪んでおり、かつ、カテゴリ数が少ない場合が挙げられる。このとき、polychoric 相関係数は母相関係数を過大推定する。また母相関係数が.96 や.98 といったように非常に高く、2 変量が逆方向に強く歪んでいるときにはカテゴリ数が多い場合に過小推定が生じる。これらのバイアスは Likert の簡便法を用いた場合よりも相対的には小さいものの、必ずしも真の相関係数を正確に反映していない可能性があるという点で注意が必要である。応用研究者がこの係数を利用する場合には、観測された分割表の周辺度数を確認し、逆方向の強い歪みが生じていないか確認してから使うことで、バイアスの小さい推定値が得られるだろう。

Polychoric 相関係数を用いない場合には、sigma 法や ridit 法の利用を検討する価値があるかもしれない。これらの手法には一定のバイアスが生じるものの、2 変量の分布が逆方向に歪んでいる場合には Likert の簡便法よりもバイアスが少なくなることが予想されるからである。双対尺度法については、カテゴリ数が増えるとともに相関係数が高くなり、母相関が低い場合には過大推定のリスクが生じることから離散化によるバイアスの対処法として利用するのは避けた方が良いだろう。

## 4.3 カテゴリ値に整数値以外を割り振った場合および polychoric 相関係数の実データによる検討 (研究 2-b)

### 4.3.1 問題と目的

シミュレーションの結果と実際のデータを比較して検討することは、手法の性質を評価するに当たって重要である。本節では、分割表の形で提供されるいくつかのデータに対して (1) Likert の簡便法, (2) sigma 法, (3) ridit 法, (4) 双対尺度法のそれぞれを用いてスコアリングした後に算出した相関係数と (5) polychoric 相関係数を算出し、それぞれの値がどのように変わるのを検討する。

### 4.3.2 方法

村瀬 (1917) において、分割表の形で報告されているデータから 2 つの表を選び (1) から (5) の手法を用いて相関係数を計算した。1 つ目の分割表は、兄弟姉妹の作文の成績についての「兄弟姉妹の綴方点数の相関表 (第 49 表)」を用いた (Table 4.3)。ここでは兄弟姉妹作文の成績がそれぞれ「丙」「乙」「甲」の 3 段階で評定されている。

Table 4.3  
兄弟姉妹の綴方点数の相関表

		兄弟			合計
		丙	乙	甲	
弟妹	丙	2	6	1	9
	乙	2	45	23	70
	甲	1	31	56	88
合計		5	82	80	167

注) 村瀬 (1917) より作成

2 つ目の分割表は「父と男子との学業成績の相関表」を用いた。このデータにおいては、父と男子の成績「3」から「10」までの 8 段階で取得されている (ただし、父



の方の成績については、「3」「4」のカテゴリは使用されなかったため、実質的には6段階と8段階の評定データである)。

Table 4.4  
父と男子との学業成績相関表

	父								合計
	3	4	5	6	7	8	9	10	
子 3	-	-	-	1	-	-	-	-	1
4	-	-	-	-	3	-	-	-	3
5	-	-	-	-	3	4	-	-	7
6	-	-	-	4	6	7	-	1	18
7	-	-	1	5	13	12	6	3	40
8	-	-	-	5	6	13	10	3	37
9	-	-	-	6	4	9	11	-	30
10	-	-	-	1	2	10	7	2	22
合計	-	-	1	22	37	55	34	9	158

注) 村瀬 (1917) より作成

### 4.3.3 結果と考察

兄弟姉妹間の綴方点数のデータに対して、各手法を用いてカテゴリに与えた得点と、その得点を用いて計算された相関係数を Table 4.5 に示した。

ここでは、Likert の簡便法、sigma 法、ridit 法、双対尺度法で得られた相関係数には大きな差はないが、polychoric 相関係数では絶対値の向上が見られる。おそらく3カテゴリという「粗い」カテゴリによる情報の損失がある程度回復されたためと思われる。

一般に実データにおいて真の値は不明であるため、いずれの手法が真の値に近いのかは分からず、polychoric 相関係数が真の相関係数を過大推定している可能性も考えられる。しかし、3章でのシミュレーション (研究1) の結果を参照すると  $\rho = .5, m = 3, n = 3$  の場合には、相関係数は .362 (一様分布条件)、.396 (等距離条件) まで低下しており、上記の数表から計算された Likert の簡便法に近い値である。

Table 4.5  
兄弟姉妹間の綴方点数へのスコアリングと相関係数

スコアリング		丙	乙	甲	相関係数
Likert 簡便法	兄姉	1	2	3	0.371
	弟妹	1	2	3	
Sigma 法	兄姉	-2.269	-0.673	0.832	0.371
	弟妹	-2.031	-0.688	0.755	
Ridit 法	兄姉	0.015	0.275	0.760	0.358
	弟妹	0.027	0.263	0.737	
双対尺度法	兄姉	-3.421	-0.696	0.927	0.376
	弟妹	-2.982	-0.640	0.814	
Polychoric 相関		-	-	-	0.501

また、4.2 節でのシミュレーション（研究 2-a）の  $\rho = .5, m = 3, n = 3$  を参照すると、Likert の簡便法は同一方向歪み（中）の条件で .400、同一方向の歪み（強）の条件で .392 であるため、おそらく兄弟姉妹の作文能力がカテゴリ化されずに連続変量として得られていたのであれば、 $r = .5$  程度であった可能性が高いと推論できるだろう。

続いて「父と男子の学業成績の相関表」に（1）から（5）までのスコアリングを施した結果を Table 4.6、相関係数を計算した結果を Table 4.7 に示した。ここでは先ほどとは異なり、Likert の簡便法、sigma 法、ridit 法、polychoric 相関係数で得られた値に大きな違いはないが、双対尺度法から得られた相関係数の絶対値が大きくなっている。

Table 4.6  
父と男子の学業成績データに各手法で割り当てた値

スコアリング		3	4	5	6	7	8	9	10
Sigma 法	父	-	-	-2.817	-1.513	-0.650	0.140	1.011	2.008
	子	-2.817	-2.172	-1.685	-1.158	-0.507	0.137	0.737	1.593
Ridit 法	父	-	-	0.003	0.076	0.263	0.554	0.835	0.972
	子	0.003	0.016	0.047	0.127	0.310	0.554	0.766	0.930
双対尺度法	父	-	-	1.750	-0.268	1.569	-0.138	-1.341	-0.078
	子	-0.681	3.982	1.506	1.029	0.689	-0.506	-0.958	-0.929

Table 4.7

父と男子の学業成績の相関表から計算される相関係数

スコアリング	相関係数
Likert 簡便法	0.264
Sigma 法	0.259
Ridit 法	0.280
双対尺度法	0.394
Polychoric 相関	0.277

相関係数の差を生み出した要因の一つには、双対尺度法のみが行列の順序の入れ替えを許容する手法であることが考えられる。双対尺度法により得られた重みを確認すると、行の重みが  $[-0.681, 3.982, 1.506, 1.029, 0.689, -0.506, -0.958, -0.929]$ 、列の重みが  $[1.750, -0.268, 1.569, -0.138, -1.341, -0.078]$  と順番の入れ替えが行われている。ここからは、そもそもカテゴリが順序尺度になっていない、データが非線形の関係を持っている等の可能性が示唆される。

研究 2-a において polychoric 相関係数が多くの条件で母相関係数を正確に推定できていたことを考慮すると、カテゴリ数が少ない場合に Likert の簡便法と Polychoric 相関係数の値の差が小さい場合は、2 変量正規分布の仮定への違反を疑う必要がある。こうした場合には polychoric 相関係数の利用は適切でないだろう。代わりに、重みにより順序の入れ替えが可能で、データ間の非線形の関係も扱うことが可能である双対尺度法の利用を検討することでデータからより多くの情報を取り出せる可能性が存在している。

実データによる解析の結果をまとめると、前節でのシミュレーションと同様に、Likert の簡便法、sigma 法、ridit 法の 3 つのスコアリング手法については、おおむね似たような値を算出した。カテゴリ数が 3 と粗いケースでは polychoric 相関係数により絶対値の向上が見られ、Likert の簡便法との差は前章でのシミュレーションに近い値が得られていたため、おそらく離散化によって失われた情報のいくらかが復元したことが示唆された。また、2 変量正規分布に従っていないと思われるデータについては、Likert の簡便法、sigma 法、ridit 法、polychoric 相関係数の値はあま

り変わらず、双対尺度法では絶対値が向上するケースが見られた。このことから、polychoric 相関係数はあくまで潜在変数が 2 変量正規分布に従っているという仮定のもとでカテゴリ化による情報損失を回復する手法であるため、その仮定が満たされない（と疑われる）場合の使用は適さないことが示唆された。

#### 4.4 本章のまとめ

本章では、カテゴリ化によって生じる相関係数のバイアスへの対処法について検討した。対処法を polychoric 相関係数を用いるものと Likert 型の項目へと最適な得点を与えることの 2 種類に大別し、各手法の性能を評価するためにモンテカルロ・シミュレーションを実施した。その結果、polychoric 相関係数が一部の条件を除き安定して真の母相関係数を推定することができることが示唆された。また、実データの解析からはデータが 2 変量正規分布に従わないことが想定される場合には polychoric 相関係数の使用は不適切で、別の手法によるデータ解析を検討すべきであることが示唆された。

## 第 5 章

# カテゴリ化が信頼性係数に与える影響

本章では、カテゴリ化による相関係数のバイアスが信頼性係数に与える影響を検討する<sup>1</sup>。信頼性係数には様々なものが提案されているため、5.1 節では心理学における信頼性係数の利用に関する先行研究を概観し、現在までに Cronbach の  $\alpha$  がもっとも広く使われていること、および、順序カテゴリデータ向けの信頼性係数も提案されているもののその検討は不十分であることを指摘する。5.2 節では、変数のカテゴリ化が  $\alpha$  へと及ぼす影響をモンテカルロ・シミュレーションによって評価する（研究 3）。続く 5.3 節では、順序カテゴリデータ向けに新しく提案されている非線形 SEM 係数の性能を評価するモンテカルロ・シミュレーションを行い（研究 4-a）、実データへの適用例を報告する（研究 4-b）。

### 5.1 信頼性係数の利用についての先行研究

心理学の研究において、測定の信頼性の問題は古くから関心を集めており、その検討の歴史は 1 世紀以上遡る (Traub, 1997)。精神測定に関する最初期の教科書である Thorndike (1904) は独立した一つの章を設けて信頼性を検討している。また、Spearman (1904) は測定誤差により希薄化した相関係数の修正のために信頼性係数

---

<sup>1</sup> 本章の内容は、小野島・椎名 (2020) および小野島・椎名 (2021) を加筆修正したものである。

を導入し、Spearman (1910) において正式に定義を与えているが、これは後の古典的テスト理論の定礎となった。信頼性を表す統計量の中でも現在までに最も広く利用されているのは Cronbach の  $\alpha$  (Cronbach, 1951) であるが、 $\alpha$  の誤解や誤用については昔から数多く指摘されており (Cortina, 1993; Schmitt, 1996)、近年では  $\alpha$  の代替案に関する議論も活発である (Sijtsma, 2009)。特に、因子分析や構造方程式モデリング (SEM) の発展に伴い、 $\omega$  係数などのモデルに基づく信頼性 (model-based reliability) が多くの心理測定学者から提案されるようになってきている (e.g., Bentler, 2009; Yang & Green, 2010)

しかしながら、 $\omega$  などの新しい信頼性係数は基本的に連続データを対象としたものである。Likert 型の項目のような順序カテゴリデータを対象とした信頼性については若干検討がなされているものの (e.g., Green & Yang, 2009; Yang & Green, 2015)、議論は十分でない。Likert 型の項目が依然として心理学におけるデータ取得の主要な手段であることや、近年の SEM を用いた研究の増加を鑑みると、SEM に基づく信頼性係数の順序カテゴリデータへの応用を検討することは経験科学としての心理学にとって重要である。

そこで本節では、まず古典的テスト理論および因子分析の観点から信頼性がどのように定義されるかを概観し、続いて順序カテゴリデータに対して提案されている新しい信頼性係数を検討する。

### 5.1.1 古典的テスト理論による信頼性

項目数が  $J$  であるテストや尺度において、受検者  $i$  の項目  $j$  における観測得点を  $X_{ij}$  とする。古典的テスト理論では項目の観測得点  $X_{ij}$  は真値  $T_{ij}$  と誤差  $E_{ij}$  の2つの部分からなるもの、すなわち

$$X_{ij} = T_{ij} + E_{ij} \quad (5.1)$$

であると考えられる。ここで  $T_{ij}$  は  $X_{ij}$  の期待値である。誤差得点  $E_{ij}$  は  $X_{ij} - T_{ij}$  でありその期待値は 0 である。また、合成得点 (個人の総得点) を  $X_i = \sum_{j=1}^J X_{ij}$  とすると、真値と誤差の合成得点はそれぞれ、 $T_i = \sum_{j=1}^J T_{ij}$  と  $E_i = \sum_{j=1}^J E_{ij}$  であ

る。古典的テスト理論は、 $E_{ij}$  は相互に独立かつ真値とも独立であると仮定し、そこから  $E_i$  の各要素は  $T_i$  と独立であるため、 $\sigma^2(\cdot)$  を分散とすると

$$\sigma^2(X_i) = \sigma^2(T_i) + \sigma^2(E_i) \quad (5.2)$$

が成り立つ。ここで信頼性は観測変数の分散の内、真値の分散が占める割合

$$\rho_X = \frac{\sigma^2(T_i)}{\sigma^2(X_i)} = 1 - \frac{\sigma^2(E_i)}{\sigma^2(X_i)} \quad (5.3)$$

として定義される。また、古典的テスト理論の枠組みでは信頼性は、 $\sigma(\cdot, \cdot)$  を共分散とするとテスト得点  $X_i$  とその平行形式のテスト得点  $\tilde{X}_i$  の相関係数として

$$\rho_{X\tilde{X}} = \frac{\sigma(X_i, \tilde{X}_i)}{\sigma(X_i)\sigma(\tilde{X}_i)} \quad (5.4)$$

としても定義される (Lord & Novick, 1968)。

歴史的には、信頼性係数の概念は誤差の混入により希薄化した相関係数を補正するために導入されたが、一度の観測データからでは信頼性が計算できないという問題 (Spearman, 1904, p.90) に対処するために、信頼性の議論は一度の観測からでも計算可能な信頼性の下界を求めること (Guttman, 1945) に移っていくこととなった (Traub, 1997)。現在広く使われている  $\alpha$  は Guttman が  $\lambda$  と表記して検討した 6 種類の信頼性の下界の内の一つ ( $\lambda_3$ ) であり、またそれ以前に Spearman-Brown の公式の一般化を目指して提案された Kuder-Richardson の公式 (Kuder & Richardson, 1937) を一般化したものである<sup>2</sup>。

$\alpha$  は

$$\alpha = \frac{J}{J-1} \left( 1 - \frac{\sum_{j=1}^J \sigma^2(X_{ij})}{\sigma^2(X_i)} \right) \quad (5.5)$$

として定義される。 $\alpha$  が信頼性の下界であるということは、すなわち

$$\rho_X = \frac{\sigma^2(T_i)}{\sigma^2(X_i)} \geq \frac{J}{J-1} \left[ 1 - \left( \frac{\sum \sigma^2(X_{ij})}{\sigma^2(X_i)} \right) \right] = \alpha \quad (5.6)$$

---

<sup>2</sup>  $\alpha$  の歴史について岡田 (2015) も参照されたい。

ということである。ここで等号は  $c_{jj'}$  を個人によらない定数として  $T_{ij} = T_{ij'} + c_{jj'}$  が成り立つとき、すなわち本質的タウ等価である場合にのみ成り立つ (Lord & Novick, 1968)。なお  $\alpha$  係数は、項目間の相関係数の平均  $\bar{r}$  と項目数  $J$  を用いて

$$\alpha = \frac{J\bar{r}}{1 + (J-1)\bar{r}} \quad (5.7)$$

としても求めることができる。このとき、Cronbach の  $\alpha$  は  $-1/(J-1) < \bar{r} \leq 1$  の区間で  $\bar{r}$  の単調増加関数である。

近年の  $\alpha$  の使用の是非に関する検討では、例えば  $\alpha$  が次元性の指標のごとく扱われるなど多くの心理学者に誤解されており誤用されていること (Cortina, 1993; Schmitt, 1996; Sijtsma, 2009)、潜在因子が 1 次元で本質的タウ等価でない場合に  $\alpha$  は信頼性を下回ること (Lord & Novick, 1968)、誤差に正の相関が生じているときに  $\alpha$  は信頼性の下界とならず上回る場合があること (Kano & Azuma, 2003) が報告されている。これらの問題を重要視し、 $\alpha$  の使用を全くやめてしまうべきであるという主張もあれば (McNeish, 2018)、 $\alpha$  が有用な場面は依然として多いとする主張もある (Raykov & Marcoulides, 2019)。これらの  $\alpha$  の利用の問題点に関する議論は、次節で紹介する因子分析モデルに基づく信頼性の利用を検討する議論へとつながった (e.g., Bentler, 2009; Yang & Green, 2010)。

### 5.1.2 因子分析モデルに基づく信頼性

因子分析モデルでは、以下のように、受検者  $i$  の  $j$  番目の項目得点である確率変数  $X_{ij}$  は切片項を 0 としたとき共通因子によって説明される部分とそうでない部分に分解される。すなわち、

$$X_{ij} = \sum_{k=1}^K \lambda_{jk} F_{ik} + \epsilon_{ij} \quad (5.8)$$

である。ここで、 $F_{ik}$  は受検者  $i$  の  $k$  番目の共通因子の得点を表す確率変数、 $\lambda_{jk}$  は  $j$  番目の項目の  $k$  番目の因子負荷量、 $\epsilon_{ij}$  は  $X_{ij}$  の内で共通因子によって説明されなかった独自因子得点である。独自因子得点は項目に固有の特殊因子と誤差因子の得



点の和からなるが繰り返しの測定なしにこれらを分離することが技術的に難しいために、合わせて扱われることが多い。また、合成得点  $X_i$  は

$$X_i = \sum_{j=1}^J \sum_{k=1}^K \lambda_{jk} F_{ik} + \sum_{j=1}^J \epsilon_{ij} \quad (5.9)$$

であり、因子分析モデルの仮定  $\forall F_{ik}, \forall \epsilon_{ij}, \sigma(F_{ik}, \epsilon_{ij}) = 0$  および  $\sigma(\epsilon_{ij}, \epsilon_{ih}) = 0 (j \neq h)$  を置けば、合成得点の分散は、

$$\sigma^2(X_i) = \sigma^2 \left( \sum_{j=1}^J \sum_{k=1}^K \lambda_{jk} F_{ik} \right) + \sigma^2 \left( \sum_{j=1}^J \epsilon_{ij} \right) \quad (5.10)$$

として表される。

因子分析に基づく信頼性係数は、(5.10) の右辺第 1 項によって表される共通因子によって説明される部分と (5.10) の右辺全体で表される合成得点の分散の比によって

$$\omega = \frac{\sigma^2(\sum_{j=1}^J \sum_{k=1}^K \lambda_{jk} F_{ik})}{\sigma^2(\sum_{j=1}^J \sum_{k=1}^K \lambda_{jk} F_{ik}) + \sigma^2(\sum_{j=1}^J \epsilon_{ij})} \quad (5.11)$$

として定義される (Raykov & Shrout, 2002)。これは McDonald の  $\omega$  (McDonald, 1985, Eq. 7.3.9) に一致する。 $\omega$  係数は (5.3) の真値の分散を、因子分析モデルの枠組における共通因子による分散に置き換えたものと考えることができる。

$\omega$  は階層的因子分析モデルの観点からも捉えることができ (Revelle & Condon, 2019)、 $\omega$  係数を算出する統計ソフトウェアでは、こちらの定義から計算されることもある。この場合には、観測変数の項目得点は  $X_{ij} = c_j g_i + \sum_{k=1}^K \lambda_{jk} F_{ik} + \epsilon_{ij}$  として分解される。ここで  $g_i$  は全ての観測値に影響を及ぼす一般因子の得点、 $F_{ik}$  は観測値のうちの一部の項目クラスターへと影響を与える  $k$  番目の群因子の得点、 $\epsilon_{ij}$  はこれらによって説明されなかった独自因子得点である。また、 $c_j$  は項目  $j$  の  $g_i$  に対する因子負荷量、 $\lambda_{jk}$  は  $j$  番目の項目の  $k$  番目の群因子に対する因子負荷量である。合成得点はここから  $X_i = \sum_{j=1}^J c_j g_i + \sum_{j=1}^J \sum_{k=1}^K \lambda_{jk} F_{ik} + \sum_{j=1}^J \epsilon_{ij}$  と書き直せる。階層的因子分析ではすべての潜在因子が直交しているという仮定を置くので、潜在因子の分散を 1 に固定したときに、標準化された合成得点の分散は、 $\sigma^2(X_i) = (\sum_{j=1}^J c_j)^2 + \sum_{k=1}^K (\sum_{j=1}^J \lambda_{jk})^2 + \sum_{j=1}^J u_j^2$  と表現される。ここで、 $u_j^2$

は項目の独自性である。これを分母にして、階層的因子分析モデルにもとづく  $\omega$  は

$$\omega = \frac{(\sum_{j=1}^J c_j)^2 + \sum_{k=1}^K (\sum_{j=1}^J \lambda_{jk})^2}{(\sum_{j=1}^J c_j)^2 + \sum_{k=1}^K (\sum_{j=1}^J \lambda_{jk})^2 + \sum_{j=1}^J u_j^2} \quad (5.12)$$

として定義される。すなわち、 $\omega$  係数は合成得点の分散の内、一般因子と群因子で説明される割合を表している。データから計算する場合は因子分析の結果得られた因子負荷量および独自性の推定値を (5.12) 式に代入して  $\omega$  の推定値が得られる。

なお、(5.12) で表した  $\omega$  係数は  $\omega_t$  と表記されることもあり、全体の分散を一般因子のみで説明する割合である  $\omega_h$  と区別して定義されている。この2つの使い分けについて、Savalei & Reise (2019) は高次因子モデルや因子間相関の高い斜交モデルである場合には前者を、双因子 (bifactor) 構造で捉えらえる場合には後者の利用を推奨している。後のシミュレーションにおいて斜交モデルを対象とするために、本章では以下  $\omega_t$  を対象とする<sup>3</sup>。

### 5.1.3 順序カテゴリ尺度に対する新しい信頼性係数

ここまで検討した信頼性係数は、すべて観測値が連続データであることを仮定していたが、観測値が順序カテゴリデータである尺度に対する新しい信頼性係数として Zumbo, Gadermann, & Zeisser (2007) が順序  $\alpha$  (ordinal  $\alpha$ ) を、Green & Yang (2009) が非線形 SEM 係数 (nonlinear SEM coefficient) を提案している。いずれも polychoric 相関係数を用いている。

Zumbo et al. (2007) は、Likert 型の項目への回答データは Pearson の相関係数の重要な前提である連続性を違反しているために歪んでおり、歪んだ相関から計算される信頼性係数も同様に歪んでいるとの認識に立ち、Polychoric 相関係数をベースとした新しい信頼性係数の順序  $\alpha$  を提案した。順序  $\alpha$  は、(5.7) 内の  $\bar{r}$  を polychoric 相関係数から計算したものに置き換えることによって算出することができる。Gadermann, Guhn, & Zumbo (2012) は、モンテカルロ法によって順序  $\alpha$  はカテゴリ数が小さいときや、カテゴリ化後の分布が歪んでいる際には、Cronbach の

<sup>3</sup>  $\omega$  係数については岡田 (2011) も参照されたい。

$\alpha$  よりも正確に信頼性を推定することを報告するとともに、R 上で動くスクリプトを提供している。

順序  $\alpha$  は提案以来その利用が広がっているが (Chalmers, 2018), その有用性を疑問視する声もある。Chalmers (2018) は、順序  $\alpha$  は従来の信頼性係数とは違って、潜在変数について仮説的な (hypothetical) 信頼性の尺度として解釈するべきで従来の信頼性係数の代替として用いることの危険性を指摘している。Savalei & Reise (2019, p.4) も同様に、polychoric 相関係数をベースにした場合にはその係数は観測された尺度得点の信頼性係数としては解釈できず、変数が仮に連続変数として取得されていた場合における合成得点に占める信頼できる分散の割合であると注意を喚起している。

Green & Yang (2009) は、順序カテゴリデータ向けの因子分析を用いて順序カテゴリデータに適した信頼性係数を提案し非線形 SEM 係数と名付けている。この係数は、順序カテゴリデータに対する因子分析 (Muthén, 1984) を実施したのちにそこで得られたカテゴリ間の閾値と因子負荷量を用いて計算される。順序カテゴリデータへの因子分析では、項目  $j$  のカテゴリ数を  $M_j$  として  $j$  番目の項目得点  $X_{ij}$  の背後に以下の関係を満たす連続量である潜在変数  $X_{ij}^*$  があると仮定する。すなわち、

$$X_{ij} = c, \quad \text{if } \tau_{j,c} \leq X_{ij}^* < \tau_{j,c+1}, \quad (c = 0, 1, \dots, M_j - 1) \quad (5.13)$$

$$(-\infty = \tau_{j,0} \quad \text{and} \quad \tau_{j,M_j} = \infty)$$

である。ここで  $\tau_{j,c}$  は項目  $j$  の  $c$  番目の閾値である。閾値と因子負荷量を用いて、非線形 SEM 係数は、次式のように定義される。

$$\rho_{nl} = \frac{\sum_{j=1}^J \sum_{j'=1}^J \left[ \sum_{c=1}^{M_j-1} \sum_{c'=1}^{M_{j'}-1} \Phi_2(\tau_{j,c}, \tau_{j',c'}, \sum_{k=1}^K \sum_{k'=1}^K \lambda_{jk} \lambda_{j'k'} \rho_{F_k F_{k'}}) - (\sum_{c=1}^{M_j-1} \Phi_1(\tau_{j,c})) (\sum_{c'=1}^{M_{j'}-1} \Phi_1(\tau_{j',c'})) \right]}{\sum_{j=1}^J \sum_{j'=1}^J \left[ \sum_{c=1}^{M_j-1} \sum_{c'=1}^{M_{j'}-1} \Phi_2(\tau_{j,c}, \tau_{j',c'}, \rho_{X_j^* X_{j'}^*}) - (\sum_{c=1}^{M_j-1} \Phi_1(\tau_{j,c})) (\sum_{c'=1}^{M_{j'}-1} \Phi_1(\tau_{j',c'})) \right]} \quad (5.14)$$

として計算される。ここで、 $\lambda_{jk}$  は  $j$  番目の潜在変数  $X_j^*$  の  $k$  番目の因子の負荷量であり、 $\rho_{F_k F_{k'}}$  は因子間相関である。 $\Phi_1(\tau_{j,c})$  は 1 変量の標準正規分布  $\phi_1(x)$  の

累積分布関数で  $\Phi_1(\tau_{j,c}) = \int_{-\infty}^{\tau_{j,c}} \phi_1(x)dx$  である。また、 $\Phi_2(\tau_{j,c}, \tau_{j',c'}, \rho_{X_j^* X_{j'}^*})$  は、非対角要素が  $\rho_{X_j^* X_{j'}^*}$  である母相関係数行列  $\Sigma$  をパラメータに持つ 2 変量の標準正規分布に従う確率密度関数  $\phi_2(x, y)$  の累積分布関数で、 $\Phi_2(\tau_{j,c}, \tau_{j',c'}, \rho_{X_j^* X_{j'}^*}) = \int_{-\infty}^{\tau_{j,c}} \int_{-\infty}^{\tau_{j',c'}} \phi_2(x, y)dydx$  である。サンプルから計算する際には、確認的因子分析の結果得られた推定値を (5.14) に代入する。

この係数は、観測得点のスケールと因子得点のスケールの違いがあるためカテゴリデータに対しては「因子による分散 ÷ 観測値の分散」による信頼性の定義が意味をなさないことを背景に、(5.4) で示した「尺度の総点とその尺度と平行な測定になるように作られた平行形式の尺度の総点の相関係数」を用いて定義された信頼性係数である。(5.4) における分子と分母は、(5.13) 内の潜在連続変数の正規性の仮定および平行形式の尺度では因子負荷量等のパラメータが等しいとする仮定を用いて式展開を行うことで (5.14) 式となる。平行形式と言っても仮説的に想定されるものであるので 1 回の測定からでも計算が可能である（詳細は Green & Yang (2009) を参照されたい）。

順序  $\alpha$  と非線形 SEM 係数はどちらも順序カテゴリデータ向けの信頼性係数であるが、それぞれの係数が「何に対する」信頼性であるかに違いがある。先に述べたように順序  $\alpha$  は潜在変数について仮説的な信頼性の尺度である一方で、非線形 SEM 係数は (5.14) の分母にあるようにカテゴリ化後の分散と共分散を用いているため、観測変数の尺度得点に対する信頼性となっている点で異なっている。したがって、順序カテゴリデータに対する従来の信頼性係数の代替として用いる場合には、順序  $\alpha$  よりも非線形 SEM 係数の方が適していると考えられる<sup>4</sup>。以下では、この係数に焦点を当てて検討を進める。

現在までに、この係数の性能を評価するためのシミュレーション研究が 2 つ行われている。一つは Yang & Green (2015) によるもので、シミュレーションで検討された大部分の条件において非線形 SEM 係数は線形 SEM 係数と同等もしくはより少ないバイアスであったこと、カテゴリ化される前の多変量の連続データにおいて正

<sup>4</sup> もちろん順序  $\alpha$  が適している場合も存在し、例えば、仮に順序カテゴリデータのカテゴリ数を増やしていった場合にどの程度まで信頼性が上昇するかを検討する場面などには使えるが、Chalmers (2018) はこうした特殊な用途以外での係数の有用性は限られていると主張している。

規性が成立しない場合においても、バイアスが少なく信頼性を推定できることが報告されている。また、Kim, Lu, & Cohen (2020) は、尺度内でカテゴリ数が異なる場合にこの係数がどのような挙動を示すかの検討を行なっている。その結果から、シミュレーションで用いられた多くの条件において、非線形 SEM 係数は  $\alpha$  に比べて信頼性係数を正確に推定すること、2 件法と 5 件法の組み合わせの場合には、全てが 2 件法の場合と全てが 5 件法の場合の間の信頼性を推定し、カテゴリ数の増加によって増加する信頼性の性質 (Masters, 1974) をよく反映している指標であると報告している。

しかしながら、非線型 SEM 係数を様々な領域で使用するためには、未検討の領域がいくつかある。1 点目は、先行研究では 1 因子モデル、双因子モデルによるものしかとりあげておらず、検討されたモデルが限定的な点である。確認的因子分析において検討されるモデルの半数以上が斜交モデルであるとする報告 (Jackson, Gillaspay, & Purc-Stephenson, 2009) を踏まえると、多因子の斜交モデルにおいてこの係数の性能を評価することが不可欠であろう。2 点目に、モデルが誤特定された場合についての評価がなされていない点である。一般にデータ解析において、真のモデルは未知であるために確認的因子分析の際に推定されるモデルは少なからず誤特定されることとなる。仮に、非線型 SEM 係数がモデルのわずかな誤特定で著しいバイアスを生じるとするのであれば、実用上の大きな問題であるため、モデルが誤特定された際の係数の挙動を調べることは応用上極めて重要である。

#### 5.1.4 本章の目的

ここまで様々な信頼性係数を概観してきたが、心理測定分野における発展が必ずしも応用研究者の信頼性係数の利用の慣習に影響を及ぼす訳ではない。2009 年に *Psychometrika* 誌上で信頼性係数の利用に関する一連の論文が掲載されたが、Sijtsma (2009) はその冒頭で  $\alpha$  のより良い代替案が存在するにも拘らずそれらはほとんど知られておらず、ましてや使われることはないとコメントしている。実際、尺度や検査において報告されている信頼性係数の頻度を調査した Hogan, Benjamin, & Brezinski (2000) は、調査対象となった尺度や検査の実に 3 分の 2 が  $\alpha$  を報告し

ていたと指摘しており、心理測定分野の発展<sup>5</sup>が研究者の信頼性係数の利用の慣習に影響を及ぼしていないことが例証されている。日本語で書かれている論文についても  $\alpha$  一辺倒の状況は同様で、 $\omega$  などの新しい信頼性係数を報告する論文はまれである (高本・服部, 2015)。新しい係数の存在自体が知られていないということも理由の一旦にあるだろうし、どのような場面で係数がうまく機能するかの性能評価が十分になされていないということも理由として挙げられるだろう。

(5.7) にあるように  $\alpha$  は相関係数の単調増加関数であるため、前章まで論じてきたカテゴリ化による相関係数の絶対値減少のバイアスは、必然的に  $\alpha$  に影響を及ぼすこととなる。変数のカテゴリ化が Cronbach の  $\alpha$  にどの程度影響を及ぼすのかを検討することは、利用される信頼性係数の多くが  $\alpha$  である現状を考慮すると重要な課題であると言えよう。5.2 節の研究 3 では、モンテカルロ法によってこの点を検討する。

また、先に非線形 SEM 係数が順序カテゴリデータ向けの信頼性係数として提案されていることを述べたが、この係数の性能評価を行った研究は少ない。特に、この係数が SEM をベースにしているものである以上、モデルの誤特定に対してどのような振る舞いをするのかを理解しておくことは、応用研究でこの係数を利用するにあたって重要であろう。5.3 節の研究 4-a では、モンテカルロ法によってこの点を検討し、5.4 節の研究 4-b において実データへの適用を報告する。

## 5.2 連続変数のカテゴリ化が Cronbach の $\alpha$ へと及ぼす影響 (研究 3)

### 5.2.1 目的

カテゴリカルなデータから計算されることによって生じた相関係数のバイアスが Cronbach の  $\alpha$  へと及ぼす影響を評価することを目的とした。

---

<sup>5</sup> Hogan et al. (2000) では心理測定分野の発展にも拘らず使われていない知識として具体的には、一般化可能性理論における一般化可能性係数や項目反応理論におけるテスト情報関数が挙げられている。

## 5.2.2 方法

シミュレーションでは、(a) 項目数、(b) 項目間の母相関係数の平均  $\bar{r}$ 、(c) カテゴリ数、(d) 「全項目数に対して床・天井効果を示す項目数の割合」(e) サンプルサイズ、の5つの要因を操作して順序カテゴリデータを生成した。生成されたデータから  $\alpha$  を計算することを1つの条件につき500回繰り返した。

### 項目数

Cronbach の  $\alpha$  が項目数の単調関数であり、項目数が大きくなると値も大きくなることはよく知られている(岡田, 2015)。ここでは一つの尺度の項目数として比較的小さいものから大きいものまでを含めるため5, 10, 20の3水準を設定した。

### 項目間の母相関係数の平均

尺度作成の論文などで報告される信頼性係数で例えば0.5より小さいなどといった低いものはあまり報告されないこと(Hogan et al., 2000)を考慮して、報告される頻度が多い.7-.9あたりの信頼性係数をカバーできるように項目間の相関係数の平均を設定した。具体的には、.3, .4, .5, .6, .7の5水準を設定した。

### カテゴリ数

質問紙調査でよく用いられるものとして、2, 3, 4, 5, 6, 7の6水準を設定した。3章や4章のシミュレーションではカテゴリ数が異なる場合に特有の挙動が見られた訳ではないので、尺度内のカテゴリ数は全て同じとした。

### 床・天井効果が項目に占める割合

連続変数をカテゴリ化する際に用いる閾値を変えることによって、床・天井効果を示す項目の割合を操作した。具体的には、床・天井効果を示す項目が全項目に対して占める割合について、0, 0.2, 0.4, 0.6の4つの水準を設定した。これは3章や4章のシミュレーションにおいて、カテゴリ化後の2変量の歪み方が相関係数のバイアスの重要な要因だと示唆されたためである。以下では「床・天井効果を示す項目」を

簡単に床・天井項目と呼ぶこととする。

連続変数をカテゴリ化する際の閾値のセットを「左右対称」、「天井効果の項目用」、「床効果の項目用」の3種類用意した。具体的には、カテゴリ数を  $M$  として左右対称の場合には  $\tau_0 = \infty, \tau_i = -2.5 + (5/n)i, \tau_M = \infty$ 、天井項目の項目では  $\tau_0 = \infty, \tau_i = -2.5 + (2.5/n)i, \tau_M = \infty$ 、床項目の項目では  $\tau_0 = \infty, \tau_i = (2.5/n)i, \tau_M = \infty$  を用いた。

床・天井項目が占める割合によって、どの項目にどの種類の閾値を用いるかを決定した。床と天井項目の個数が偶数となる場合には、床効果項目と天井効果項目の個数は同一とした。例えば、項目数が10で床・天井効果の項目の割合が0.4の場合には、床効果の項目用の閾値で2つの変数を、天井効果の項目用の閾値で2つの変数を、残りの6つの項目を左右対称の閾値でカテゴリ化した。床効果と天井効果の項目数が奇数の場合には、以下のようにした：(1) 項目数が5で床・天井効果の割合が0.2の場合は、1つの項目を天井効果の閾値を用い、床効果の閾値は用いない。(2) 項目数が5で床・天井効果の割合が0.6の場合は、1つの項目に床効果の閾値を用い、2つの項目に天井効果の閾値を用いた。

## サンプルサイズ

50, 100, 500 の3条件を設定した。

## データ生成の手続き

期待値ベクトル  $\boldsymbol{\mu} = \mathbf{0}$  と母相関行列  $\Sigma_{m \times m} = (\sigma_{ij} = 1 \quad (i = j); \sigma_{ij} = \bar{r} \quad (i \neq j))$  をパラメータに持つ多変量正規分布に従う乱数を発生させた後に、決められた割合の項目を床・天井効果になる閾値のセットで、それ以外の項目を左右対象な閾値セットでカテゴリ化した。カテゴリ化の際に全ての値が同じカテゴリに属してしまい分散が0の項目が生じる場合には再度サンプリングをした。カテゴリ化後の  $\alpha$  を記録し、条件のセル毎に500回ずつ繰り返した。



## ソフトウェア

シミュレーションは全て統計解析環境 R version 3.6.3(R Core Team, 2020)で行った。追加のパッケージとして、 $\alpha$ の算出には psych version 1.9.12(Revelle & Condon, 2019), 多変量データの生成には mvtnorm version 1.1-0(Genz et al., 2020)に所収の関数を使用した。

### 5.2.3 結果と考察

Figure 5.1 に、すべてのサンプルサイズを合算したデータにおいて、条件ごとに Cronbach の  $\alpha$  を平均したものを示した。ここからは、全体的な傾向として (1) 項目間母相関の平均が高いほど、カテゴリ化によって  $\alpha$  が大きく低下すること、(2) 床・天井効果の項目が多く含まれるほど  $\alpha$  が大きく低下すること、(3) カテゴリ数が少ないほどカテゴリ化によって  $\alpha$  が大きく低下すること、(4) 項目数が少ない場合に、カテゴリ化によって  $\alpha$  は大きく低下すること、の 4 点を指摘することができる。この傾向はサンプルサイズ毎に分けて平均を算出した場合でも同様であった。サンプルサイズ毎の平均のグラフは付録 C に示してある。

カテゴリ化後の  $\alpha$  と (5.7) に母相関係数の平均を代入して算出される期待値との差を検討すると、項目数が多くなるほど、またカテゴリ数が多くなるほどこの差は縮まるが、この程度は床・天井効果の項目が含まれる割合によって大きく変わっていた。例えば、項目数が 10、カテゴリ数が 4、 $\bar{r} = .5$  の条件では、連続変数で  $\alpha$  を算出した際の期待値が .909 であるが、カテゴリ化後の  $\alpha$  の平均は、床・天井項目が含まれない場合には .883、2 割が床・天井項目の場合には .869、4 割が床・天井項目の場合には .854、6 割が床・天井項目の場合には .837 と、床・天井効果を含む項目の割合が大きくなるとともに差が大きくなる傾向が見られた。

続いて  $\alpha$  のばらつきを見るために、条件ごとに  $\alpha$  の SD を算出したものを Figure 5.2 に示した。ここからは (1) サンプルサイズが大きいほど  $\alpha$  のばらつきは小さくなる傾向があること、(2) 項目数が多いほど  $\alpha$  のばらつきは小さくなる傾向があること、(3) カテゴリ化の際のカテゴリ数が大きくなるほど  $\alpha$  のばらつきは小さくな

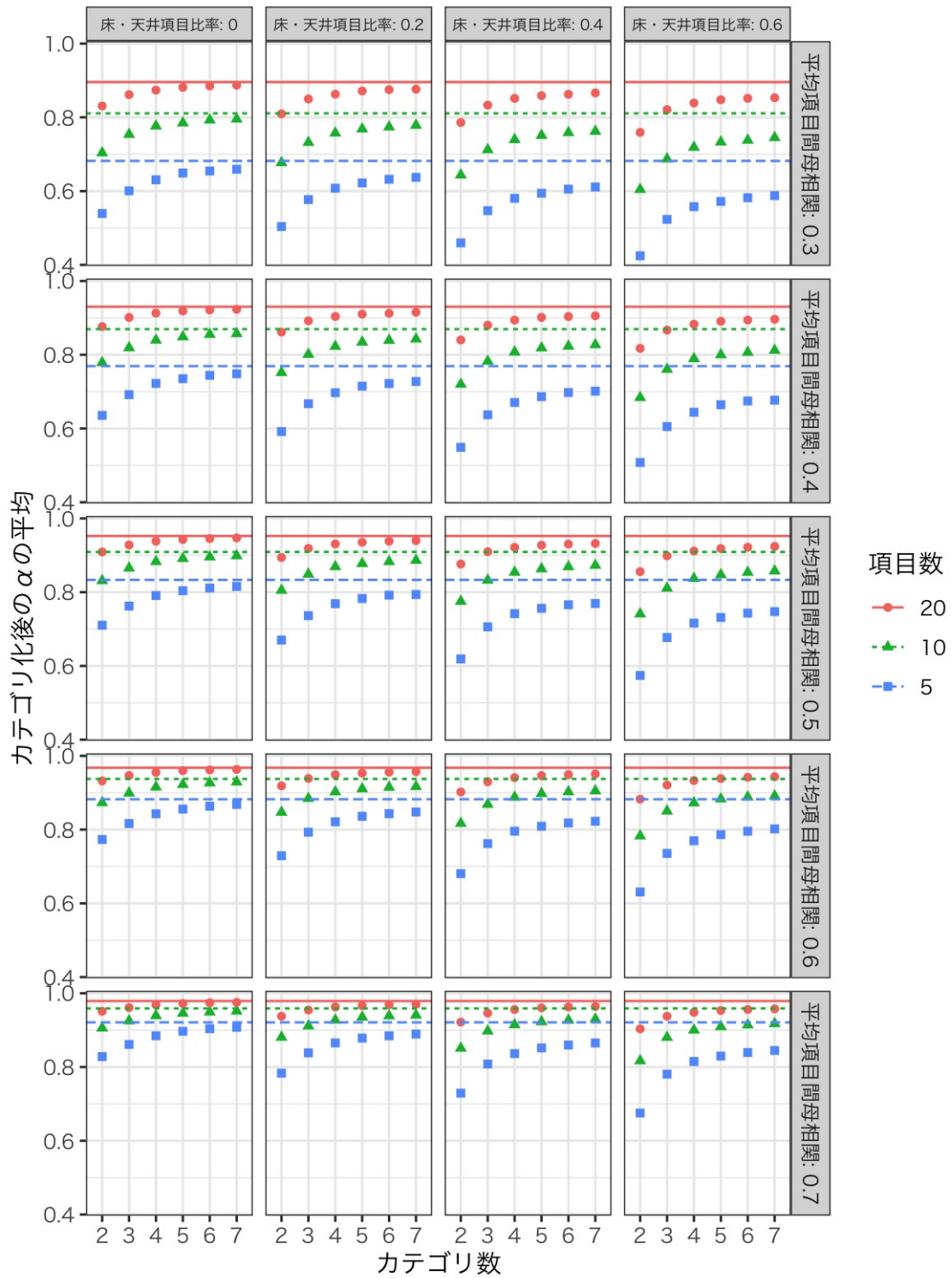


Figure 5.1. カテゴリ化後の  $\alpha$  の平均

注) 水平線は (5.7) に項目間母相関の平均と項目数を代入した際の期待値

る傾向があること、(4) 床・天井効果が含まれる項目の割合が小さいほど  $\alpha$  のばらつきは小さくなる傾向があること、の 4 点が指摘できる。これらの傾向は  $\bar{r}$  の値にかかわらず見られた。

研究 3 の結果は、変数のカテゴリ化によってカテゴリ数が少ない場合に信頼性係数には大きなバイアスが生じるが、カテゴリ数が多くなるとそのバイアスが小さくなるという Masters (1974) の報告と一貫している。さらに、カテゴリ数を増やしたとしても、床・天井効果を示すような項目が多い場合には一定のバイアスが生じうるという新たな知見も示した。

床・天井効果が含まれる割合が大きいほどカテゴリ化によって大きく  $\alpha$  が低下したことは、3 章や 4 章のシミュレーションの結果と整合する。すなわち、床・天井効果を生み出すような閾値のもとでは、もとの多変量正規分布の構造が破壊されることになるため、カテゴリ化後の相関係数も大きく低下することになる。また、2 変量が逆方向に歪んでいる場合には相関係数はより顕著に低下する。これらの情報の損失が  $\alpha$  の低下を招くことに繋がったと考えられる。

本研究の結果から尺度作成などの応用場面における留意点を述べる。Cronbach の  $\alpha$  が項目数が増えると高くなることや項目間の相関係数が高い場合に高くなることはよく知られており、尺度に含む項目数を決定する際に参考にされている。しかしながら、本研究の結果から、尺度に含まれる項目に床効果や天井効果のような歪みが含まれる場合には信頼性は思うように高くなるということが示唆されるため、予備調査などを実施して項目の中に一定の歪んだ分布を示すものが存在する場合には、項目数をより多く見積もるなどの必要があろう。また、歪んだ項目が多い場合にはカテゴリ化による相関係数の情報損失は大きくなるため、次節で扱う非線形 SEM 係数のように潜在変数の情報損失を考慮した手法を用いることを検討する必要があるだろう。

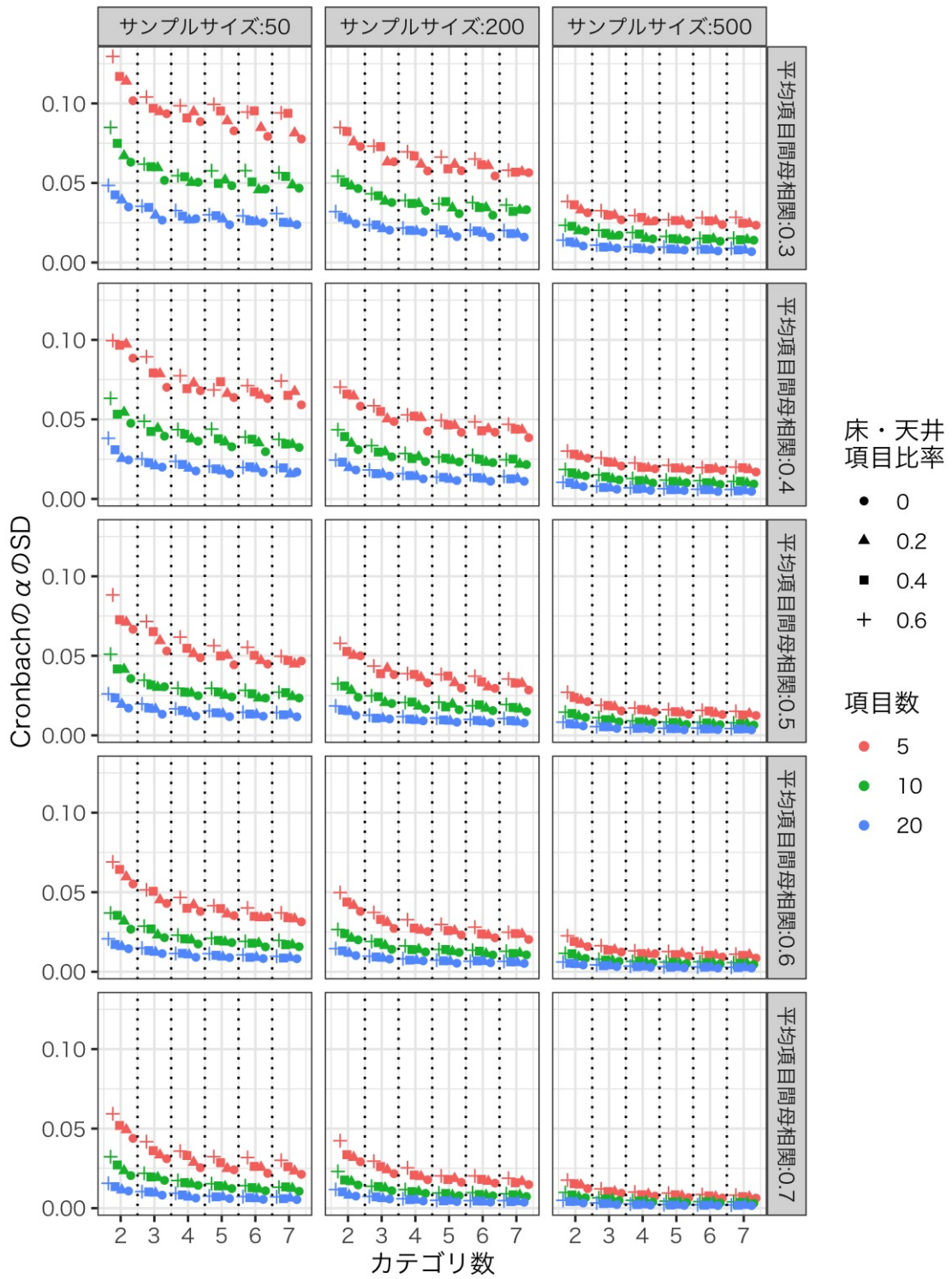


Figure 5.2. カテゴリ化後の  $\alpha$  の SD

## 5.3 順序カテゴリデータに対する因子分析に基づく信頼性係数の評価（研究 4-a）

### 5.3.1 目的

多因子構造のモデルに対して、モデルが誤特定された場合において非線型 SEM 係数がどのような挙動を示すかを検討することを目的とした。

### 5.3.2 方法

Yang & Green (2015) を参考に、シミュレーションは 2 段階に分けて行った。本来であれば、信頼性係数の期待値を数式変形のみで求め、その値との差をバイアスとするべきであるが、特に (5.13) のような変換が含まれる場合に、これは非常に難しい。そこで、1 段階目の研究として、まずは十分に大きいサイズのデータを発生させ、それを母集団の近似とみなす「近似母集団研究」によるシミュレーションを実施し、そこで得られた値を信頼性係数の期待値とみなすこととした。続いて、2 段階目の研究として応用研究における現実的なサンプルサイズを設定し、サンプリングを繰り返す「サンプルデータ研究」によるシミュレーションを実施し、近似母集団研究で得られた期待値との差を検討した。それぞれの段階での、要因および水準の設定には Kim et al. (2020) による条件も参考にした。近似母集団研究では以下の 3 つのデザイン因子を操作した：(a) 因子構造（モデル）、(b) カテゴリ数、(c) 項目の閾値の均質性。サンプルデータ研究ではこれらに加えて (d) サンプルサイズの要因も操作し、各要因が信頼性係数へと与える影響を検討することとした。

#### 因子構造（モデル）

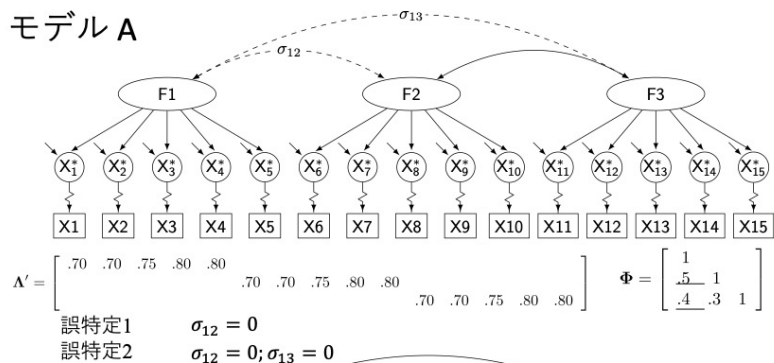
シミュレーションで扱うモデルは実際の因子分析の使用例と近いものが望ましいため、Jackson et al. (2009) による確認的因子分析の結果報告に関するレビューを参考に全部で 3 つのモデルを設定した。Jackson et al. (2009) は、文献検索の結果から確認的因子分析を用いて報告されるモデルにおいて、因子数の中央値は 3 (90 %

近くが6因子より少ないこと), 50% 近くのモデルが1次の斜交モデル(次いで多いのが1次の直交モデルで12%), 観測変数の数の中央値は17(25パーセンタイルが12, 75パーセンタイルが24)であることを報告している。この結果および, モデルの誤特定と適合度の関係を検討した Hu & Bentler (1998) や Sivo, Fan, Witta, & Willse (2006) で用いられたものをベースにして, 典型的な因子数・項目数のモデルとしてシンプルな測定モデルと交差負荷を含むやや複雑なモデルを検討した。モデル A は, 潜在的な次元は3次元で因子一つあたりの観測変数は5である斜交モデルとした (Figure 5.3 A)。モデル B は, モデル A を基本としながらも, いくつかの項目で交差負荷があるモデルとした (Figure 5.3 B)。また, 因子数が少ないモデルは Yang & Green (2015) で扱われているものの, 因子数や観測変数の数が大きいモデルについては未検討であるため, Jackson et al. (2009) のサーベイを参考に3つ目のモデルとして因子数が6, 観測変数が24であるモデル C も検討することにした (Figure 5.3 C)。確認的因子分析では, 特定の項目群について, 項目間の内容の類似性, 測定方法の類似性, 時間的な類似性など共通因子では説明されない共通の変動要因を持つことがあり, それは誤差の共分散としてモデルに組み込まれることがあるため (狩野, 2002), モデル C では誤差共分散を含めた。それぞれのモデルの因子負荷量行列および因子間相関は Figure 5.3 に示した。

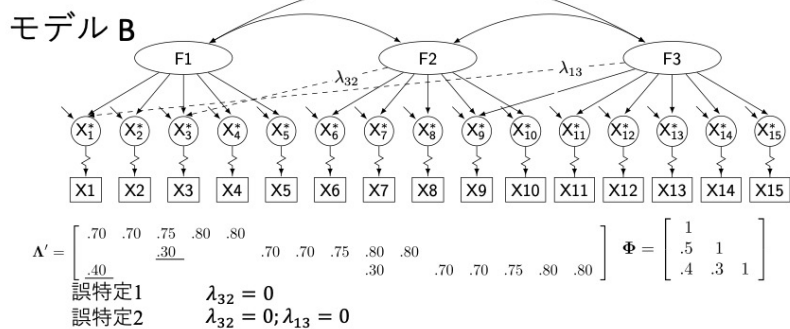
## カテゴリ数

カテゴリ数は2, 3, 4, 5, 6の5水準を設定した。Kim et al. (2020)の研究では, 異なるカテゴリ数の組み合わせに特有な効果は見られなかったことを考慮し, カテゴリ数は尺度間で全て同じにした。

### モデル A



### モデル B



### モデル C

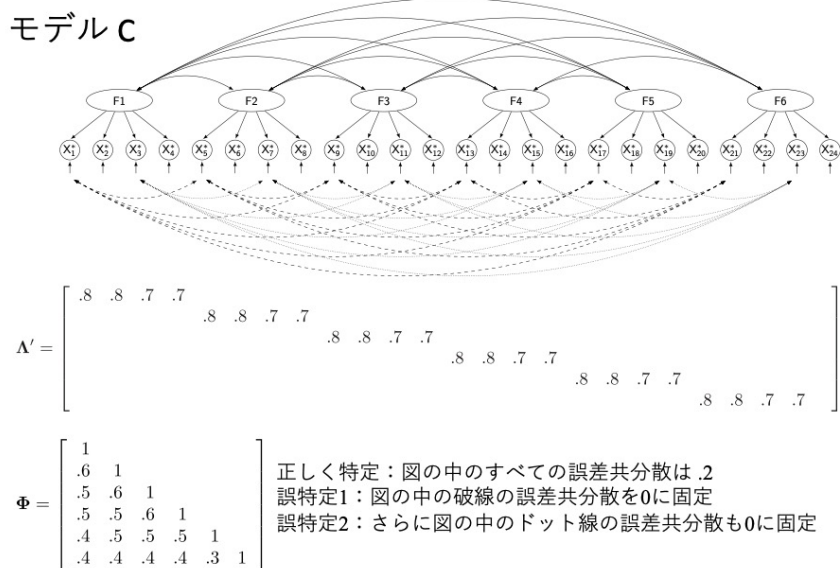


Figure 5.3. シミュレーションで用いられたモデル

注 1) モデル A, B で下線が引かれている数値は誤特定で 0 に固定されるパラメータ。

注 2) モデル C の観測得点は紙面の都合で省略してある。

## カテゴリ化の際の閾値と閾値の均質性

すべての項目は (5.13) を適用してカテゴリ化した。カテゴリ化の際の閾値  $\tau_{j,c}$  は Yang & Green (2015) を参考にして、項目のカテゴリ数を  $M_j$  としたときに  $M_j - 1$  個の閾値が「左右対称」, 「主に負」, 「主に正」に分布する 3 種類のものを設定した。それぞれの閾値の具体的な値は Table 5.1 に示した。これらの値は、閾値が「左右対称」の場合には、カテゴリ化後の分布が左右対称の単峰になるように、「主に負」, 「主に正」については単調に増えていく（減っていく）分布になるように、カテゴリ化後の分布を目視して確認しながら探索的に決定した。これらの閾値を用いて、Yang & Green (2015) と同様に閾値の均質性について以下の 4 つの条件を設定した：(a) 全ての項目が「左右対称」の閾値でカテゴリ化される「均質 (homogeneous)」条件、(b) 項目の一部が「主に負」の閾値で、項目の一部が「主に正」の閾値でカテゴリ化され、残りは「左右対称」の閾値でカテゴリ化される「中程度に不均質 (moderately heterogeneous)」条件、(c) 項目の半数近くが「主に負」と「主に正」の閾値で、残りが「左右対称」の閾値でカテゴリ化される「かなり不均質 (highly heterogeneous)」条件、(d) 奇数番目の項目を -1 の閾値で、偶数番目の項目を 1.5 の閾値でカテゴリ化する「極端に不均質 (extremely heterogeneous)」条件 (d 条件はカテゴリ数が 2 の場合のみ)。各条件でどの項目にどの閾値が用いられたかを Table 5.1 の注に示した。

Table 5.1  
シミュレーションで用いた閾値

カテゴリ数	閾値の種類		
	主に負	左右対称	主に正
2	-0.5	0	0.5
3	-1, 0	-0.75, 0.75	0, 1
4	-1.2, -0.6, 0	-0.8, 0, 0.8	0, 0.6, 1.2
5	-1.5, -1, -0.6, 0	-1.2, -0.5, 0.5, 1.2	0, 0.6, 1, 1.5
6	-1.8, -1.2, -0.7, -0.3, 0.25	-1.75, -0.8, 0, 0.8, 1.75	-0.25, 0.3, 0.7, 1.2, 1.8

注) 「中程度に不均質」の条件では、モデル A およびモデル B において項目 1, 6, 11 に主に負の閾値が用いられ、項目 2, 7, 12 に主に正の閾値が用いられた。モデル C において項目 1, 5, 9, 13, 17, 21 に主に負の閾値が用いられ、項目 2, 6, 10, 14, 18, 22 に主に正の閾値が用いられた。「かなり不均質」の条件では、モデル A およびモデル B において項目 1, 3, 6, 8, 11, 13 に主に負の閾値が、項目 2, 4, 7, 9, 12, 14 に主に正の閾値が用いられた。モデル C において項目 1, 3, 6, 9, 11, 14, 17, 19, 22 に主に負の閾値が、項目 2, 5, 7, 10, 13, 15, 18, 21, 23 に主に正の閾値が用いられた。



## データ発生の手続き（近似母集団研究）

Yang & Green (2015) の手続きを参考にして、 $n = 1,000,000$  のデータを発生させてそれを母集団の近似とした。正規分布に従う多変量の乱数は  $\mu=0$  と  $\Sigma$  をパラメータに持つ多変量正規分布から発生させた。ここで  $\Sigma = \Lambda\Phi^t\Lambda + \Psi$  であり、真のモデルによって構造化された相関係数行列である。 $\Lambda$  はモデルの因子負荷量行列、 $\Phi$  は因子間相関行列、 $\Psi$  は独自性を対角成分に持つ行列である。モデルごとの  $\Lambda$  と  $\Phi$  は Figure 5.3 に示した。diag( $\cdot$ ) を行列の対角成分を列ベクトルに変換する操作としたときに、 $\Psi$  はモデル A とモデル B においては  $\mathbf{1} - \text{diag}(\Lambda\Phi^t\Lambda)$  を対角要素に持つ対角行列、モデル C においてはそれに Figure 5.3 内に示した誤差共分散を非対角要素に加えたものである。生成された多変量の連続データに対して、(5.13) を適用しカテゴリ化して順序カテゴリデータを生成した。

## モデルの誤特定と非線形 SEM 係数の計算

生成された順序カテゴリデータに対して、確認的因子分析を実施し、非線形 SEM 係数の計算に必要なパラメータを得た。その際に、誤特定の程度を操作し、同一データに対して (a) モデルが正しく特定された場合 (correctly specified), (b) モデルの一部の係数を 0 に固定して誤特定した場合 (misspecified 1), (c) モデル内のより多くの係数を 0 に固定して誤特定した場合 (misspecified 2) の 3 つの場合で確認的因子分析を行った。それぞれのモデルにおいて具体的にどの係数が 0 に固定されたかは Figure 5.3 に示した。モデルの誤特定については、多様なものが考えられるが、モデルの誤特定とモデルに基づく信頼性係数との関係は先行研究もなく不明な点も多いため、まずはモデルの誤特定の研究においてよく用いられる、母集団における 0 でないパラメータを 0 に固定する誤特定に焦点を絞ることとした。モデルの識別のために潜在変数の分散はすべて 1 に固定した。推定量には順序カテゴリデータでカテゴリ数が小さい場合の確認的因子分析においてしばしば推奨される平均と分散で検定統計量を調整する重み付き最小二乗法 (WLSMV)<sup>6</sup>を指定した (Finney &

<sup>6</sup> WLSMV という用語は構造方程式モデリングのソフトウェアの発展の中で定着した独特の用語であり、これは推定量のみを指定している訳ではない。実際 version 6 以降の Mplus や本研究で用い

DiStefano, 2013)。その後、得られた  $\tau_{j,c}, \lambda_{jk}, \rho_{X_j^* X_j^*}, \rho_{F_k F_k'}$  を用いて (5.14) から非線形 SEM 係数を算出した。

また、Yang & Green (2015) を参考に (5.4) の平行測定による信頼性係数の計算のために、項目得点についての平行形式のデータ ( $n = 1,000,000$ ) を作成した。まず、(a) 因子間相関  $\Phi$  をパラメータに持つ多変量標準正規分布に従う因子得点ベクトル  $F$  と (b)  $\mu=0$ , 分散共分散行列が  $\Psi$  の多変量正規分布に従う 2 組の独自因子得点ベクトル得点  $R^{*1}$  と  $R^{*2}$  を生成し、

$$X^{*1} = \Lambda F + R^{*1}, \quad X^{*2} = \Lambda F + R^{*2} \quad (5.15)$$

を適用して、項目得点についての 2 組の連続データを生成した。

この 2 組のデータに (5.13) を用いてカテゴリ化したのちに、それぞれの合成得点間の相関係数を (5.4) 式より計算し、平行測定による信頼性係数の期待値の近似値  $\rho_{X\bar{X}}$  とした。

また、応用場面では依然として  $\alpha$  の利用が主流であること (高本・服部, 2015),  $\alpha$  係数の代替案としてしばしば  $\omega$  係数の利用が推奨されること (Revelle & Condon, 2019; 清水, 2007) を踏まえて、(5.5) と (5.12) を用いて  $\alpha$  と  $\omega$  を算出した。なお、 $\omega$  係数の算出に必要な因子負荷量の推定値は、探索的因子分析を用いて高次因子モデルの解を得たのちに Schmid-Leiman 変換 (Schmid & Leiman, 1957) を用いて得る方法と、確認的因子分析を用いて双因子モデルの解を直接得る方法があるが、前者を用いた。高次因子モデルの識別のための群因子の数はモデル A, B では 3 を、モデル C では 6 を指定した。確認的因子分析ではなく探索的因子分析による手法を用いた理由は、R のパッケージ psych などに実装されている関数ではこちらの手法がデフォルトとして用いられており、応用研究者が用いることが多いだろう手法を比較対象として設定したかったからである。

モデルに基づく信頼性の評価においては、データへの当てはまりも重要な要素であるため (cf. Savalei & Reise, 2019), 確認的因子分析の際の適合度指標 (CFI,

---

た R のパッケージである lavaan では、推定量に対角成分のみ重みづける対角重み付き最小二乗法 (DWLS) を指定し、scaled かつ shifted な検定統計量を用いる方法を意味しており、平均と分散 (MV) で調整した検定統計量を用いてはいない。この用語については Rosseel (2014) が詳しい。

RMSEA) も合わせて取得した。

### データ発生の手続き (サンプルデータ研究)

サンプルデータ研究においては、近似母集団の研究と同じ要因に加えて、サンプルサイズの要因を操作した。サンプルサイズの水準は 50, 100, 200, 500, 1000 の 5 水準であった。1 つの条件につき 500 回の繰り返しを行った。近似母集団と同様の手続きで連続データを生成しカテゴリ化をした後に、(5.14) を用いてモデルが正しく特定された場合の非線形 SEM 係数、モデルが誤特定された場合の非線形 SEM 係数 2 種類、 $\alpha$  係数、 $\omega$  係数を算出した。適合度も近似母集団の研究と同様に記録した。なお、カテゴリ化の際に分散が 0 になる変数が生じた場合には再度サンプリングを行った。

### ソフトウェア

シミュレーションは全て統計解析環境 R version 3.6.3 (R Core Team, 2020) で行った。追加のパッケージとして、 $\alpha$  および  $\omega$  の算出には psych version 1.9.12 (Revelle & Condon, 2019)、多変量データの生成および、非線形 SEM 係数の算出の際の二変量正規分布の累積確率の計算には mvtnorm version 1.1-0 (Genz et al., 2020)、確認的因子分析には lavaan version 0.6-5 (Rosseel, 2012) に所収の関数を使用した。非線形 SEM 係数については Yang がホームページで公開している SAS のコードおよび Kim et al. (2020) がオンライン資料として公開している物をもとに関数を定義した。

### 5.3.3 結果

まず近似母集団研究の結果を示した後に、サンプルデータ研究の結果を示す。近似母集団において (5.4) から計算した平行測定による信頼性を  $\rho_{X\bar{X}}$  と表記する。また、(5.14) から計算した非線形 SEM 係数を指すときは、モデルが正しく特定されている場合は  $\rho_{correct}$ 、モデルの一部が誤特定されている場合 (misspecified 1) は  $\rho_{mis1}$ 、モデルのより多くの部分が誤特定されている場合 (misspecified 2) は  $\rho_{mis2}$  と表記する。サンプルデータにおいても同様に、 $r_{correct}$ 、 $r_{mis1}$ 、 $r_{mis2}$  と表記する。また

同じ条件における  $\rho_{X\tilde{X}}$  との差をバイアスと定義し、 $b_{correct}$  や  $b_{mis1}$  というように表記する。例えば、サンプルデータで、モデル B、カテゴリ数 2、閾値が均質の条件で、正しく特定された非線形 SEM 係数  $r_{correct}$  の場合は、同じ条件下（モデル B、カテゴリ数 2、閾値が均質）の  $\rho_{X\tilde{X}}$  との差  $r_{correct} - \rho_{X\tilde{X}}$  を  $b_{correct}$  とする。本節では、主要な結果のみを示し条件ごとの詳細な数表は付録 D にて示す。

### 近似母集団における信頼性係数

すべての条件で分散が負の値をとるなどの不適解は発生しなかった。Figure 5.4 および Table D 1 に  $\rho_{correct}, \rho_{mis1}, \rho_{mis2}, \alpha, \omega$  および、平行測定による信頼性係数の期待値の近似値である  $\rho_{X\tilde{X}}$  を示した。Figure 5.4 および Table D 1 からは以下 6 点が指摘できる。(1) モデルが正しく特定されている場合には全ての条件で  $|b_{correct}| < .010$  であり、 $\rho_{correct}$  は  $\rho_{X\tilde{X}}$  に極めて近い値であった。(2) モデルが誤特定された場合には、非線形 SEM 係数には  $\rho_{X\tilde{X}}$  との間にバイアスが生じた。バイアスの大きさおよび符号はモデルの種類によって異なり、モデル A、モデル C、モデル B の順番に大きかった。モデル A では  $\rho_{mis2} < \rho_{mis1} < \rho_{correct} \simeq \rho_{X\tilde{X}}$  であり、 $\rho_{X\tilde{X}}$  を下回った一方で、モデル B およびモデル C では、 $\rho_{mis2} > \rho_{mis1} > \rho_{correct} \simeq \rho_{X\tilde{X}}$  と、 $\rho_{X\tilde{X}}$  を上回る傾向があった。いずれのモデルでも誤特定の程度が大きくなると  $\rho_{X\tilde{X}}$  との差も大きくなる傾向が見られた (i.e.,  $|b_{mis2}| > |b_{mis1}|$ )。 (3) カテゴリ数と誤特定によるバイアスの程度に明確な関係は見られなかった。(4) いずれのモデルでも、項目が不均質になるほど、信頼性係数は低下する傾向が見られた。非線形 SEM 係数においては、閾値の均質性とバイアスの程度に明確な関連性は見られなかった。(5)  $\alpha$  係数については、モデル A、モデル B では一貫して過小推定、モデル C では過大推定した。(6)  $\omega$  係数は、モデル A、モデル B においては項目が極端に不均質な場合を除いて真の信頼性係数に近い値であった。誤差相関を含むモデル C では、過大推定する傾向があった。

それぞれのモデルにおける適合度 (CFI と RMSEA) を Table 5.2 に示した。予想されたことではあるが、条件ごとのそれぞれのデータに対して「正しく特定」されたモデルがもっともよい適合を示し、次いで「誤特定 1」のモデルが良い適合を示

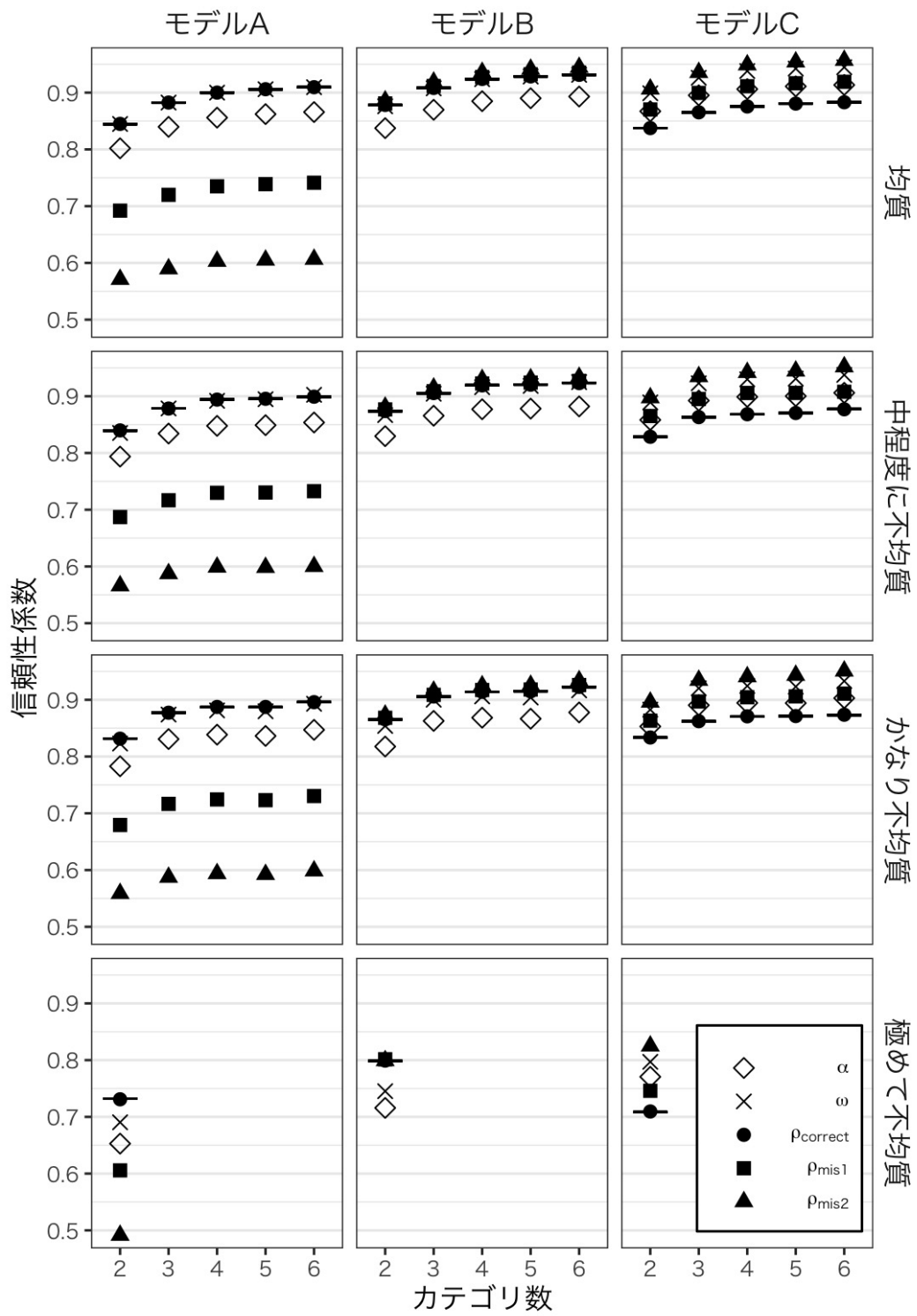


Figure 5.4. 近似母集団における信頼性係数

し、「誤特定 2」のモデルがもっとも悪い適合度であった。また、誤特定による非線形 SEM 係数のバイアスが大きかったモデル A では、誤特定による適合度の悪化も大きく、誤特定による非線形 SEM 係数のバイアスが小さかったモデル B では誤特定による適合度の悪化も小さかった。これらのことから、誤特定によって生じる非線形 SEM のバイアスは適合度と関連していると考えられた。

Table 5.2  
 近似母集団研究における CFI（上段）／RMSEA（下段）

閾値の均質性	モデル A			モデル B			モデル C		
	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2
2 カテゴリ									
均質	1.00	.841	.788	1.00	.992	.973	1.00	.971	.941
	.000	.099	.113	.000	.027	.050	.000	.037	.050
中程度に不均質	1.00	.843	.788	1.00	.992	.975	1.00	.971	.944
	.000	.096	.111	.000	.026	.047	.000	.035	.048
かなり不均質	1.00	.843	.786	1.00	.993	.976	1.00	.975	.949
	.000	.093	.109	.000	.024	.045	.000	.033	.045
極めて不均質	1.00	.850	.764	1.00	.996	.982	1.00	.982	.946
	.001	.055	.068	.000	.013	.026	.000	.019	.032
3 カテゴリ									
均質	1.00	.834	.788	1.00	.991	.968	1.00	.964	.928
	.000	.121	.135	.001	.036	.065	.000	.048	.065
中程度に不均質	1.00	.836	.789	1.00	.991	.969	1.00	.963	.928
	.000	.119	.134	.000	.035	.064	.000	.047	.064
かなり不均質	1.00	.838	.790	1.00	.991	.969	1.00	.966	.931
	.000	.116	.131	.000	.034	.063	.000	.045	.062
4 カテゴリ									
均質	1.00	.834	.794	1.00	.989	.963	1.00	.958	.916
	.001	.129	.143	.001	.041	.075	.000	.054	.074
中程度に不均質	1.00	.836	.795	1.00	.990	.966	1.00	.958	.920
	.000	.126	.140	.000	.039	.070	.000	.052	.069
かなり不均質	1.00	.837	.792	1.00	.991	.967	1.00	.964	.928
	.000	.122	.137	.000	.036	.067	.000	.048	.065
5 カテゴリ									
均質	1.00	.834	.796	1.00	.989	.962	1.00	.957	.913
	.000	.134	.148	.000	.043	.079	.000	.057	.077
中程度に不均質	1.00	.837	.797	1.00	.989	.966	1.00	.955	.918
	.000	.130	.144	.000	.041	.073	.000	.055	.072
かなり不均質	1.00	.836	.793	1.00	.991	.968	1.00	.963	.927
	.000	.125	.139	.000	.037	.068	.000	.049	.067
6 カテゴリ									
均質	1.00	.834	.797	1.00	.989	.961	1.00	.956	.911
	.000	.137	.151	.000	.045	.082	.000	.058	.080
中程度に不均質	1.00	.835	.798	1.00	.989	.964	1.00	.954	.913
	.000	.134	.148	.000	.043	.078	.000	.057	.076
かなり不均質	1.00	.835	.796	1.00	.990	.964	1.00	.960	.920
	.000	.130	.144	.000	.041	.075	.000	.053	.073

## サンプルデータにおける信頼性係数

サンプルデータについては 500 回の繰り返しの平均を用いて評価した。非線形 SEM 係数および  $\omega$  を算出する際の因子分析の結果で分散が負の値をとるなど不適解が生じた場合には計算から除外した。

Table 5.3  
サンプルデータにおける不適解発生率

サンプルサイズ	モデル A				モデル B				モデル C			
	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$
2 カテゴリ												
50	49.9%	50.2%	47.1%	0.4%	73.2%	74.6%	70.4%	0.4%	100.0%	100.0%	100.0%	0.8%
100	15.2%	16.1%	19.5%	0.3%	44.3%	49.4%	38.6%	0.2%	50.0%	44.0%	53.8%	1.5%
200	1.4%	2.2%	3.6%	0.1%	16.4%	29.8%	13.9%	0.1%	10.8%	13.9%	24.9%	1.5%
500	0.6%	0.6%	0.8%	0.0%	4.0%	19.5%	4.0%	0.0%	1.6%	3.1%	2.1%	0.1%
1000	0.4%	0.4%	0.4%	0.0%	1.3%	13.7%	2.5%	0.0%	0.5%	1.1%	0.2%	0.0%
3 カテゴリ												
50	12.9%	11.7%	10.0%	0.0%	100.0%	99.9%	49.1%	0.1%	100.0%	100.0%	100.0%	0.4%
100	0.5%	0.3%	0.4%	0.0%	16.7%	34.2%	16.9%	0.0%	100.0%	100.0%	12.3%	0.1%
200	0.0%	0.0%	0.0%	0.0%	2.7%	20.3%	3.5%	0.0%	0.1%	0.1%	1.1%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	7.5%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4 カテゴリ												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.4%
100	0.1%	0.3%	0.1%	0.0%	9.4%	23.7%	9.0%	0.0%	100.0%	100.0%	100.0%	0.1%
200	0.0%	0.0%	0.0%	0.0%	1.0%	12.9%	0.8%	0.0%	0.1%	0.0%	0.4%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	3.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5 カテゴリ												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.1%	100.0%	100.0%	100.0%	0.6%
100	0.1%	0.0%	0.0%	0.0%	7.4%	22.9%	7.0%	0.0%	100.0%	100.0%	100.0%	0.1%
200	0.0%	0.0%	0.0%	0.0%	1.1%	11.8%	0.8%	0.0%	0.0%	0.0%	0.6%	0.1%
500	0.0%	0.0%	0.0%	0.0%	0.0%	2.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6 カテゴリ												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.4%
100	9.9%	9.9%	9.9%	0.0%	19.4%	31.9%	19.5%	0.0%	100.0%	100.0%	100.0%	0.1%
200	0.0%	0.0%	0.0%	0.0%	0.2%	9.3%	0.2%	0.0%	0.0%	0.0%	0.4%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

非線形 SEM 係数の不適解の発生率はデータ発生条件によって大きく異なった。項目の均質性の条件を考慮しないで平均したものを Table 5.3 に示した。項目の均質性の水準毎のそれぞれの詳細な不適解の発生率は付録の補足資料に示した (Table D 2-6)。全体的な傾向として、(1) モデルが複雑なほど、すなわち推定するパラメータ数が多いモデルほど不適解発生率が高く、(2) サンプルサイズが小さいほど不適解の発生率が高かった。また (3) サンプルサイズが 200 を超えると、大部分の条件

で不適解発生率は 10% 以下になったが、カテゴリ数が 2 の場合のモデル B、モデル C、カテゴリ数が 3 以上の場合のモデル B の誤特定 1 では 10% を超える不適解が発生した。 $\omega$  係数については、大部分の条件で 1% 以下の不適解の発生率であり、最も高くともモデル C のサンプルサイズ 100, 200 の 1.5% であった。

各条件の信頼性係数の平均および比較のための  $\rho_{X\bar{X}}$  を Figure 5.5 に示した（「極端に不均質」の条件グラフの判読性を考慮し省略した）。非線形 SEM 係数は近似母集団での結果と同様に（1）誤特定の程度が大きいほどバイアスが大きくなること、（2）バイアスの符号はモデルによって異なることを示した。また（3）サンプルサイズが小さい場合には、仮にモデルが正しく特定されていたとしても、信頼性係数を過大推定する傾向があることが確認された。誤特定によるバイアスの程度と閾値の均質性の間には明確な関係は見られなかった。数表は付録の Table D 7-9 に示した。

$\alpha$  係数はモデル A, B においては  $\rho_{X\bar{X}}$  を過小推定する傾向に、モデル C においては過大推定する傾向にあった。 $\omega$  係数は、近似母集団での結果と同様にモデル A および B では、大部分の条件で  $|b_\omega| < .010$  と  $\rho_{X\bar{X}}$  に近い値であったが、カテゴリ数が 2 の場合で項目が著しく不均一な場合にはモデル A では  $-.020$  から  $-.040$  程度、モデル B では  $-.035$  から  $-.052$  程度の過小推定をしていた。モデル C では過大推定の傾向が見られ、その程度は  $.051$  から  $.146$  程度であった。このバイアスは  $r_{correct}$  とは異なり、サンプルサイズの大きさに関わらず生じた。

各サンプルサイズでの RMSEA と CFI の平均を付録の Table D 10-14 に示した。全体的な傾向としていずれのモデルも、誤特定の程度が深刻なほど RMSEA は大きくなり、CFI は小さくなる傾向が見られた。また、適合度の悪化の程度はモデルの種類によって異なった。モデル A では、誤特定により適合度が大きく低下し、モデル C では中程度悪化し、モデル B では適合度は大きくは変わらなかった。近似母集団での研究と同様に、適合度の悪化の程度は非線形 SEM 係数のバイアスの程度を反映していた。



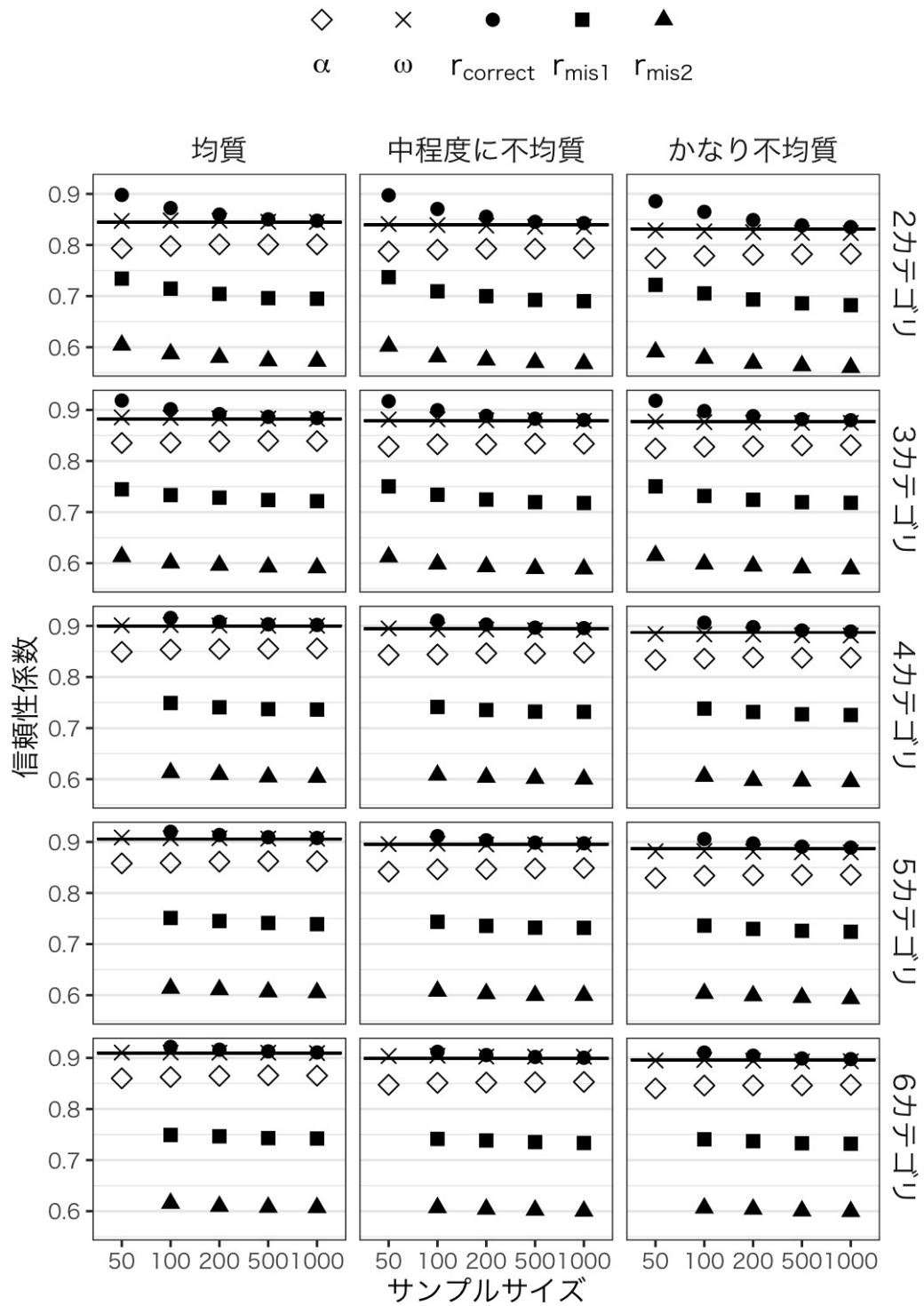
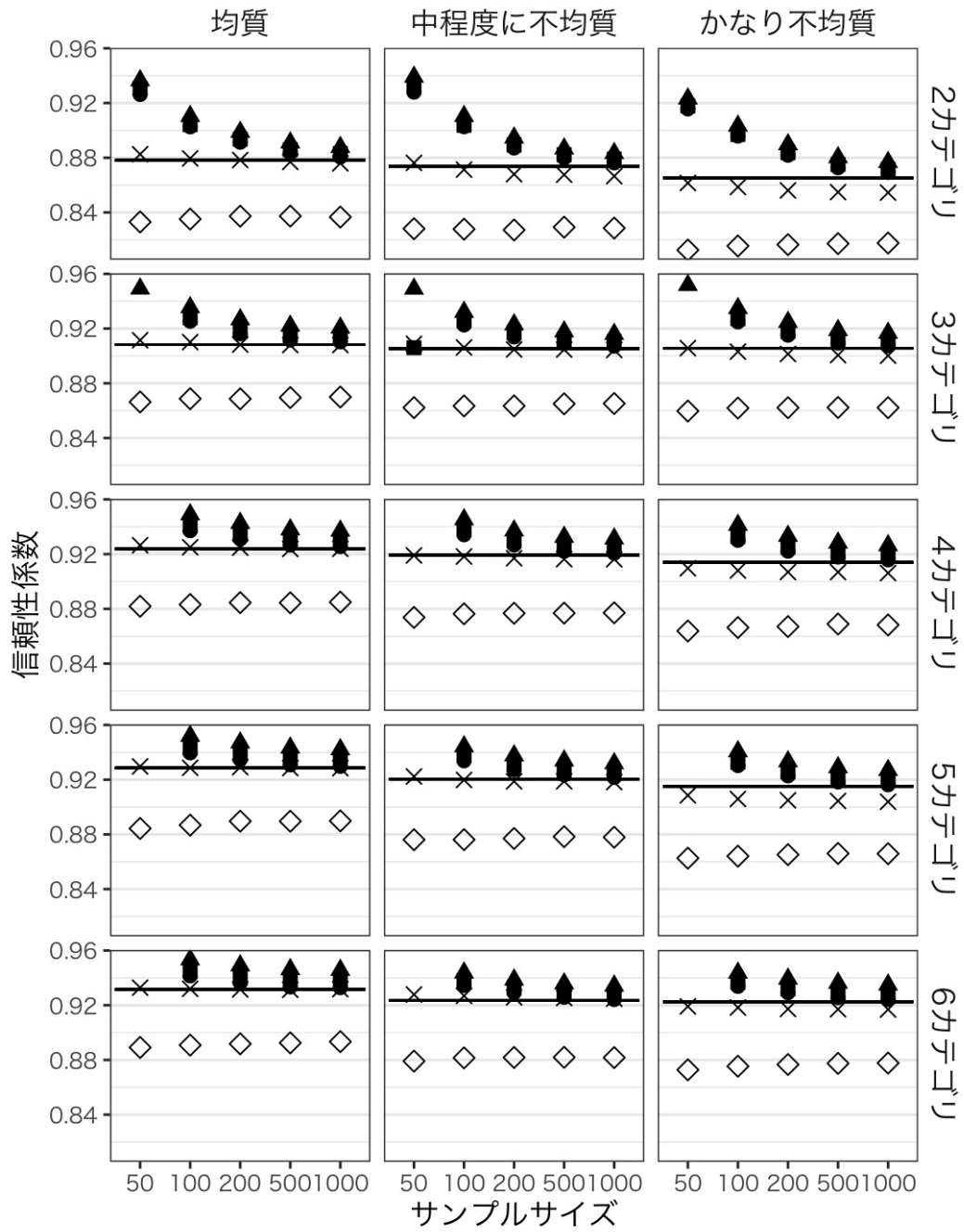


Figure 5.5. サンプル研究の平均信頼性係数

$\diamond$     $\times$     $\bullet$     $\blacksquare$     $\blacktriangle$   
 $\alpha$     $\omega$     $r_{\text{correct}}$     $r_{\text{mis1}}$     $r_{\text{mis2}}$



(b) モデル B

Figure 5.5. サンプル研究の平均信頼性係数 (続き)

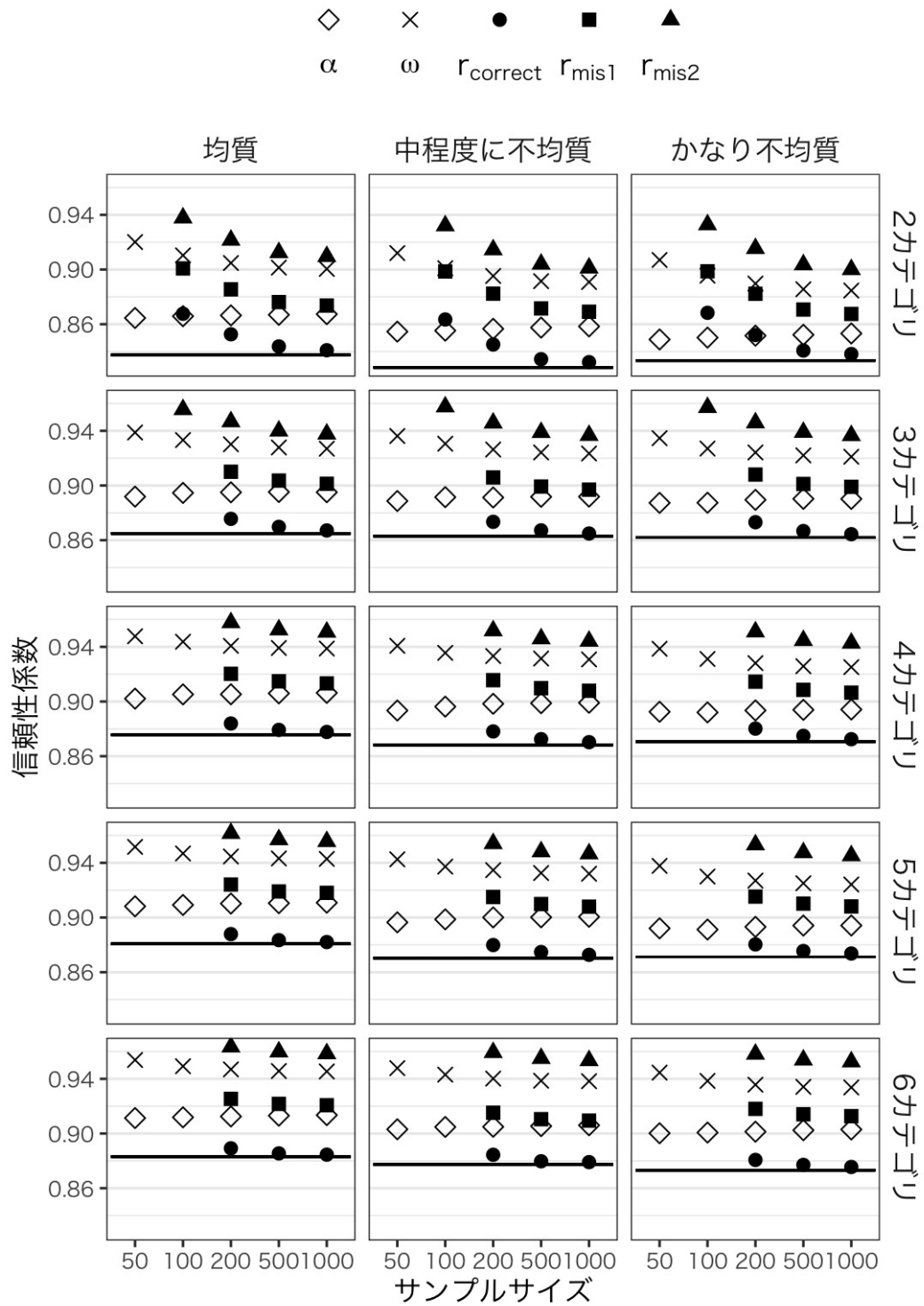


Figure 5.5. サンプル研究の平均信頼性係数 (続き)

### 5.3.4 考察

研究 4-a では、順序カテゴリデータ向けの信頼性係数である非線形 SEM 係数の推定の正確性を、モデルが誤特定された状況下に焦点を当てて評価した。シミュレーションの結果、非線形 SEM 係数はモデルが正しく特定されている限りにおいては極めて正確な推定が可能だが、モデルが誤って特定された場合には、深刻なバイアスが生じる可能性があることが明らかになった。バイアスの程度は、モデルの種類やモデルの誤特定の程度に依存した。

この結果は Yang & Green (2015) や Kim et al. (2020) による報告と整合している。すなわち、先行研究で検討された潜在次元が次元の場合や階層的構造（双因子モデル）の場合に加えて、本研究の結果は非線形 SEM 係数が斜交モデルの多次元構造においても正確な推定が可能であり、多様な場面で活用可能なことを示したものである。

また、本研究はモデルが誤って特定された場合には深刻なバイアスが生じうるという新たな知見も示した。モデルの誤特定によるバイアスの程度と、モデルの適合度は関連していた。すなわち、適合度が悪くなるほど非線形 SEM 係数のバイアスも大きくなった。これが意味することは、非線形 SEM 係数は「モデル」に基づく信頼性係数であるため、そのもととなるモデルがうまくデータを説明できていない場合には、信頼性を正確に推定できないということである。

加えて本研究はサンプルサイズが小さい場合には不適解が発生しそもそも非線形 SEM 係数が計算できない場合があること、仮に計算ができた場合でも信頼性係数を過大推定する傾向があることも示した。このことは、非線形 SEM 係数を計算しようとする場合には、モデルのサイズに応じて十分に大きいサンプルサイズを用意する必要があることを示している。

非線形 SEM 係数はモデルが正しく特定されており、かつ、十分なサンプルサイズである場合には  $\omega$  係数と比較して同等あるいはより高い正確性を示したが、この差の解釈には注意が必要である。方法の節でも述べたことだが、本シミュレーションにおける  $\omega$  は探索的な手法によるものを用いているので、非線形 SEM 係数と  $\omega$  の差

には、カテゴリ化に由来する差とモデルの誤特定に由来する差が混じっている。モデル C において  $\omega$  がいずれの条件でも過大推定をしているのはカテゴリ化による影響よりも、誤差共分散を無視した影響が強いだろうことから、この差は主にモデルの誤特定に由来すると考えることができるだろう。その一方で、モデル A の極めて不均質な条件において非線形 SEM 係数と  $\omega$  係数に生じている差は、主にカテゴリ化によって生じているものと考えるのが妥当であろう。正しく特定されたモデルにおいて非線形 SEM 係数が閾値の均質性からの影響を受けなかったのは、順序カテゴリデータの因子分析では polychoric 相関係数の利用によりカテゴリ化による情報損失の回復が行われているからだと考えられる。他方で、 $\omega$  が極めて不均質な閾値の条件下で信頼性を過小推定したことはカテゴリ化によって損失した情報が大きく、それが因子負荷量にバイアスを生じさせたからだと推察できる。この結果は Yang & Green (2015) におけるシミュレーションの結果とも一致し、また第 3 章のシミュレーションにおいて「整っていない閾値」では相関係数のバイアスが大きかった事実とも整合的である。

研究 4-a の結果から、応用研究者が尺度作成やその後の研究での尺度利用の場面で非線形 SEM 係数を含むモデルに基づく信頼性係数の報告を行う際の留意点をいくつか述べる。第 1 に、信頼性係数の報告の際にはもととなるモデルがデータに当てはまっているかを検討することが最も肝要であろう。非線形 SEM 係数はシミュレーションの結果が示すように、モデルが誤って特定された場合には大きなバイアスが生じる危険性がある。Savalei & Reise (2019) が警鐘を鳴らすように、モデルに基づく信頼性を利用する際には、その係数の算出に用いられるパラメータを得たモデルそのものを十分に注意深く検討しなければならない。シミュレーションの結果からは、バイアスの程度はモデルの誤特定と関連しており、誤特定の程度はモデルの適合度によって検討できることが示唆されるため、適合度を吟味し注意深い潜在変数のモデリングを行なうことが望ましい。応用研究者が SRMR, CFI, RMSEA などといった適合度指標をもとにモデルの当てはまりを検討する時には、単一の指標に頼るのではなく、Hu & Bentler (1999) が推奨するように複数の異なる性質の適合度指標を組み合わせて総合的な判断を行うとともに、モデルが既存の理論と整合的で

あるかを確認することが求められるだろう<sup>7</sup>。統計ソフトウェアの進歩で複雑な計算もコンピュータが自動的に算出してくれる現在、モデルに基づく信頼性を計算する前の検討は特に重要であろう。 $\alpha$  係数の報告が、「みそぎ (岡田, 2015, p.80)」あるいは通過儀礼のような使われ方をしてきた過去を考慮すると、モデルに基づく信頼性係数も同じ運命を辿る危険性はある。上述したような注意深い検討がなされずに報告される信頼性係数は信頼できないものとなるだろう。

第2に、非線形 SEM 係数を計算する際のサンプルサイズである。シミュレーションの結果が示すように、小さいサンプルサイズでは不適解が発生し係数の計算ができなかったり、モデルが正しく特定されていた場合においても、過大推定の危険があったりするため、小さいサンプルサイズで非線形 SEM 係数を利用することは推奨できないだろう。必要なサンプルサイズはモデルの大きさやカテゴリ数に依存するため、一律の基準を設けることは難しいが、シンプルなモデルにおいては 200、複雑なモデルにおいては 500 程度のサンプルサイズが最低でも必要になるだろう。

留意点の3点目は、他の信頼性係数との使い分けに関してである。まず、本質的タウ等価が仮定できない場合や次元尺度でない場合に  $\alpha$  は信頼性を一貫して過小推定するので多くの心理測定学者が主張するように (e.g., Sijtsma, 2009),  $\alpha$  のみに頼った報告はやめるべきであろう。 $\omega$  係数については、信頼性を酷く過小推定することはないが、カテゴリ数が少なく、項目が著しく不均一な場合には過小推定することもあり、このバイアスはサンプルサイズが大きくなったとしても改善はしないので、非線形 SEM 係数との併用が望ましいであろう。また、真のモデルが誤差共分散を含むときには探索的因子分析を用いて算出される  $\omega$  係数では、過大推定のバイアスが生じるため、この場合にも非線形 SEM 係数との併用が望ましいだろう (あるいは  $\omega$  を計算する際には誤差共分散を指定した確認的因子分析の解を用いて計算する必要があるだろう)。Revelle & Condon (2019, p.1408) は、信頼性係数の報告において、今後は少なくとも2つ以上を報告することが望ましいと主張しているが、分析するデータが順序カテゴリデータである場合は、非線形 SEM 係数は有望な選択肢になる

---

<sup>7</sup> Hu & Bentler (1999) は SEM における適合度指標の報告の際の基準として最も引用される文献の一つだが、Marsh, Hau, & Wen (2004) のようにその一般化可能性に疑問を呈する声も存在している。

だろう。

本研究のシミュレーションにはいくつかの課題もある。1点目は、シミュレーションで扱ったモデルの誤特定の種類が限定的な点である。本研究で、3通りのパスの削除による誤特定に焦点を当てたが、これは数あるモデルの誤特定の一部でしかない。より多様な種類の誤特定について条件を追加した検討が必要だろう。また、種類のみでなく、誤特定の程度を操作してバイアスを評価することも有益だろう。Fan & Sivo (2007) は、モデルの適合度が誤特定の種類の違いを反映するのかを検討するシミュレーションを行っており、その中で非心分布のパラメータと自由度から計算される検定力を用いて誤特定の程度を数量化している。こうした指標を用いて、どの程度の深刻な誤特定が非線形 SEM 係数のバイアスにつながるのかを検討することは有益であろう。2点目は、分布の仮定への違反について扱うことができていない点である。例えば、潜在連続変数や誤差の分布の正規性への違反がある場合にこの係数がどのような挙動を示すのかを検討することは有益であろう<sup>8</sup>。3点目は、現実に用いられている多様なモデルを十分に含めることができていない点である。今回シミュレーションで扱ったモデルは、因子数や項目数において確認的因子分析で用いられる典型的なモデルであるとは言えるが、例えば、因子負荷量の大きさなどは操作されていない。より多様なモデルでのシミュレーションによる検討が望まれる。

## 5.4 順序カテゴリデータに対する因子分析に基づく信頼性係数の実データへの適用（研究 4-b）

### 5.4.1 目的

前節のシミュレーションで扱った各信頼性係数が実際のデータにおいてどのような挙動を示すかを検討することは、係数の理解を深めるために有用である。そこで、本研究では、Rosenberg 自尊感情尺度の回答データを用いて、モデル特定の条件を

---

<sup>8</sup> Yang & Green (2015) は潜在連続変数の正規性もシミュレーションの中で操作しており、非線形 SEM 係数への影響は少ないと結論づけているが、近年、シミュレーション内で使われている非正規の乱数生成の手法には問題点があり正規性の違反への評価になっていないとの指摘がなされており (Grønneberg & Foldnes, 2019)、その結果については慎重に評価する必要がある。

変えた非線形 SEM 係数,  $\alpha$  係数,  $\omega$  係数の比較を目的とした。

## 5.4.2 方法

### データ

R のパッケージ Lambda4 version 3.0(Hunt, 2013) に所収の Rosenberg 自尊感情尺度の回答データを用いた。回答は高校生とカレッジの学生 857 名を対象としたものである。

### モデル

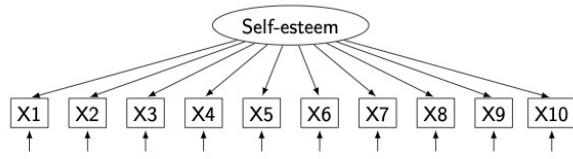
Rosenberg (1965) の自尊感情尺度は 4 件法の Likert 式の 10 項目から構成され、自尊感情と同方向の 5 項目、逆方向の 5 項目から構成される<sup>9</sup>。理論的には 1 次元を想定して作られたものだが、探索的因子分析では 2 因子以上を報告する研究も多く、自尊感情が 1 因子からなるのか、項目表現別の 2 つの自尊感情からなるのか、1 因子ではあるが同一方向の項目表現間に共分散が残っているためなのか、その因子構造について現在までに様々な測定モデルが提案されている (清水・吉田, 2008)。

ここでは (1) 自尊感情の 1 因子を想定したシンプルなモデル、(2) 否定表現の項目 (オリジナル版の pp.17-18 における 2, 5, 6, 8, 9 番目の項目) の独自因子間に共分散を設定したモデル、(3) 肯定表現と否定表現の 2 因子による 3 種類のモデルを確認的因子分析で当てはめた (Figure 5.6)。モデルの識別のために因子の分散は全て 1 に固定した。研究 4-a のシミュレーションのモデル C では、本来存在する誤差共分散をないものとして当てはめたモデルから計算された非線形 SEM 係数には過大推定のバイアスが生じた。仮に研究 4-a と同じように誤差共分散があるモデルが真のモデルであった場合に、研究 4-b におけるモデル 1 とモデル 2 はそれぞれ「誤特定」されたモデルと「正しく特定」されたモデルに対応することとなる<sup>10</sup>。

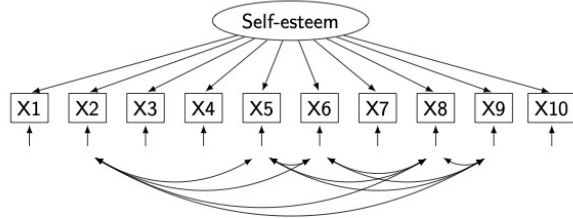
<sup>9</sup> Rosenberg (1965) の自尊感情尺度はもとは Guttman 尺度として構成されたが、現在では Guttman の方法によるスコアリングが行われることはほぼないと思われるため、本稿では Likert 法によるスコアリングによって分析を進める。

<sup>10</sup> もちろん真のモデルは未知であるため、どちらが正しく特定されたモデルに近いかを事前に知ることができない。仮に誤差共分散がないモデルが真のモデルであった場合には、研究 4-b のモデル 1 は「正しく特定」されたモデルで、モデル 2 は本来存在しないパスを追加してしまう「誤特定」をしたモデルであると言える。

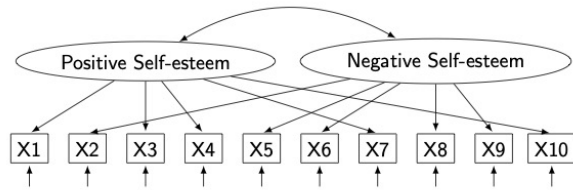




(a) 自尊感情の1因子を想定したシンプルなモデル



(b) 否定表現の項目の独自因子間に共分散



(c) 肯定表現と否定表現の2因子のモデル

Figure 5.6. 研究 4-b で当てはめた 3 つのモデル

当てはめたそれぞれのモデルのパラメータの推定値を利用して非線形 SEM 係数を計算した。 $\omega$  係数に関しては、群因子の数が 2 以下ではモデルの識別ができないため探索的因子分析ではなく確認的因子分析を用いて、上記の 3 つのモデルをあてはめそこで得られた因子負荷量から算出した。研究 4-a で用いられた探索的手法による  $\omega$  とは異なり、それぞれのモデルを直接指定している点で異なっていることに注意が必要である。

### 5.4.3 結果と考察

Table 5.4 に Rosenberg 自尊感情尺度への回答データへ 3 つのモデルを当てはめた結果およびそこから計算される信頼性係数を示した。独自因子の共分散を想定しないモデル 1 や肯定表現、否定表現の 2 因子構造を想定するモデル 3 は、誤差共分散

Table 5.4

自尊感情尺度の質問紙データへの各モデルの適合度と信頼性係数

モデル	$\chi^2$	<i>df</i>	CFI	RMSEA	SRMR	<i>r</i>	$\omega$	$\alpha$
モデル 1 (1 因子)	1230.451	35	.913	.202	.102	.932	.883	.880
モデル 2 (誤差共分散)	441.434	25	.970	.141	.049	.747	.737	.880
モデル 3 (2 因子)	625.219	34	.957	.144	.061	.921	.907	.880

注) *r* は非線形 SEM 係数を表す。

散を設定したモデル 2 に比べて非線形 SEM 係数、 $\omega$  係数共に信頼性係数が高かった。適合度については、モデル 2 が最も良い値を示した。いずれのモデルでも非線形 SEM 係数は  $\omega$  係数より若干高い値を示した。

実データの分析からは、前節のシミュレーションにおけるモデル C の結果が示唆するように、誤差共分散を設定したモデルは誤差共分散を設定しないモデルと比較して非線形 SEM 係数が低いことを示した。もっとも、実データの分析では真のモデルを知ることはできないため、どのモデルが真のモデルに近いものかを判断することはできないが、適合度の観点から、仮に検討したモデルの中ではモデル 2 が最も真のモデルに近いとした場合、モデル 1 は誤特定の程度がモデル 2 より大きいこととなり、シミュレーションの結果と合わせて考えると、モデル 1 は信頼性係数を過大推定している可能性が高いと推論できる<sup>11</sup>。

Table 5.4 の同一モデルの中で非線形 SEM 係数と  $\omega$  を比較すると、シミュレーションのモデル C に見られたほどの大きな差はなかった。この差が生じた一因として、シミュレーションでは  $\omega$  の算出にあたって探索的因子分析を用いた一方で、実データでは確認的因子分析を用いたことが挙げられる。探索的因子分析によるモデルに基づく信頼性係数はよほど大きいサンプルサイズでない限りは過大推定のバイアスが生じることが指摘されている (Savalei & Reise, 2019)。また、誤差共分散の影響をコントロールすることはできないため、モデル C のような場合には必然的に

<sup>11</sup> モデル 2 の RMSEA については Hu & Bentler (1999) で示されるカットオフの基準と比較して十分な値を示していないためさらなるモデルの改善の余地はあるが、Rosenberg 自尊感情尺度の測定モデルの探求が本研究の主たる目的ではないため、ここではこれ以上検討しない。

過大推定をすることとなり、本研究のような実データとシミュレーションの差を生んだと考えられる。

研究 4-b の結果から示唆されることは、非線形 SEM 係数はモデルに基づく信頼性係数であるため、モデルが異れば値が大きく異なるというシミュレーションの結果は実データにおいても当てはまるということである。したがって、応用研究者がこの係数を利用する際には、研究 4-a の考察部（5.3.4 項）で述べたような注意深いモデルの検討が必要となるだろう。

## 5.5 本章のまとめ

本章では、変数のカテゴリ化によるバイアスが信頼性係数へと与える影響を検討した。5.2 節のモンテカルロ・シミュレーションでは、離散化前の  $\alpha$  と離散化後の  $\alpha$  を比較した。その結果、信頼性係数は離散化の際のカテゴリ数が大きくなるほどバイアスが小さくなるものの、床・天井項目が多く含まれる場合には依然として大きなバイアスが生じることが明らかとなり、離散化に伴う情報損失への対処の必要性が示唆された。5.3 節では、モデルが誤特定された状況下における非線形 SEM 係数の挙動を検討した。その結果、非線形 SEM 係数はモデルが正しく特定されている状況下においては、項目が極めて不均質な閾値を用いてカテゴリ化された場合でも信頼性の正確な推定が可能であるが、モデルが誤特定された状況下では深刻なバイアスが生じることを明らかにし、この係数を応用研究において利用する場合のモデルを注意深く検討することの重要性を指摘した。5.4 節の実データの分析からもシミュレーションでの結果と同様にモデルが異れば非線形 SEM 係数は大きく異なることを示し、モデル検討の重要性を再度指摘した。

## 第 6 章

# 結論と展望

### 6.1 各章のまとめ

本研究の目的は、順序カテゴリデータから計算される相関係数に生じるバイアスとその対処法の評価を行うこと、および、順序カテゴリデータから計算される相関係数のバイアスが信頼性係数へと及ぼす影響の評価とその対処法の検討であった。これらの目的を達成するために、一連のモンテカルロ・シミュレーションを実施した。

まず、第 2 章において相関係数のバイアスの数理的性質を検討した。その結果、カテゴリ化後の相関係数の期待値は変数の離散化の際に用いられる閾値の関数であることを指摘した。相関係数に生じるバイアスにおいてカテゴリ数は大きな影響を及ぼすが、バイアスは本質的には離散化によってもとの分布の構造が破壊されることによって生じており、仮にカテゴリ数が大きい場合においても相関係数が大きく下がりを示すことを例示した。その上で、バイアスの評価および対処法の評価においては閾値の要因を含めた検討が必要であることを主張した。

続いて第 3 章の研究 1 において、従来のシミュレーションでは検討されてこなかった変数の離散化の際の閾値を操作して、相関係数のバイアスを評価するシミュレーションを実施した。その結果、閾値が多様な値をとりうる場合は、閾値が等距離の場合と比較して大きなバイアスを生じることを報告した。この結果は従来のシミュレーション研究は Likert 型の項目から得られる順序カテゴリデータから計算される相関係数のバイアスを過小評価していることを示唆しており、バイアスへの対処を

工夫しないとデータから誤った結論を導く可能性があることを示唆していると考えられる。

続く第4章の研究2-aではバイアスへの対処法を検討するためのシミュレーションを実施した。Likertの簡便法、sigma法、ridit法、双対尺度法を用いて各カテゴリに数値を割り当てた上で相関係数を計算する方法と、分割表からpolychoric相関係数を用いて計算する方法を比較した結果、バイアスへの対処法としてはpolychoric相関係数が優れていることを報告した。また、研究2-bでは実データに対してこれらの手法を用いた結果を報告し、polychoric相関係数は2変量正規分布に従う連続変数が離散化された場合には有効に機能するものの、そうでないことが推定される場合にはデータの変数間の構造をうまく捉えられない可能性を指摘した。

第5章では相関係数のバイアスが信頼性係数へと及ぼす影響を検討した。研究3では、現在もっとも広く用いられているCronbachの $\alpha$ へと及ぼす影響を評価した。その結果、離散化の際のカテゴリ数が多いほど $\alpha$ に生じるバイアスは小さくなるという従来の知見に加えて、変数の中に床・天井効果を示す項目が多い場合には、カテゴリ数が多くなったとしても $\alpha$ には一定のバイアスが生じうることを示した。また、研究4-aでは順序カテゴリデータ向けの信頼性係数として提案されている非線形SEM係数の性能を評価するシミュレーションを実施した。その結果、非線形SEM係数はモデルが正しく特定されている条件下では、項目が極めて不均質な閾値でカテゴリ化されるような場合でも正確に信頼性を推定できるが、モデルが誤特定された場合には大きなバイアスが生じうることを明らかにした。研究4-bでは実データへの適用例を報告し、この係数を用いる際のモデルの注意深い検討の重要性を確認した。

## 6.2 本研究の意義

相関係数は変数間の関係を記述する基本的な道具の一つである。であるからこそ、3.1節で論じたように、その基本的な道具にどのような場面でバイアスが生じるのかについては昔から検討がなされてきた。本研究は、それらの研究の流れの中に位置づけるものであるが、本研究の特色として離散化の際の閾値の要因を検討している点が

挙げられる。閾値の要因は、研究 1 においては離散化の際の閾値の生成のアルゴリズムとして、研究 2-a においては離散化後のカテゴリ分布の歪みの程度および 2 変量の歪みの方向として、研究 3 においては項目群に含まれる床・天井効果を示す項目の割合として、研究 4-a では閾値の均質性として、それぞれのシミュレーションで操作されている。

心理学を含む社会科学・行動科学においては、データ取得は多様な状況でなされる。すなわち閾値モデルのもとで、全ての参加者が全ての項目に全ての状況で同一の閾値を用いていると考えるよりは、異なる参加者、異なる項目、異なる状況ではいずれも異なる閾値が用いられると考えた方が、データ取得の状況の多様性を反映していると考えられるだろう。2.2 節で論じたように、用いられる閾値によってはカテゴリ化後の分布は、カテゴリ化前の分布の構造を保存せず相関係数に大きなバイアスをもたらす。したがって、本研究で報告された結果は、現実のデータ取得の多様性に即した相関係数のバイアスの評価になっていると考えられる。

相関係数を用いた研究は心理学において昔から広く用いられている (Rucci & Tweney, 1980)。近年の心理学において用いられるデータ分析の手続きをレビューした Blanca, Alarcón, & Bono (2018) によれば、変数間の関係を検討する指標の中でもっとも使われているものは Pearson の相関係数である。また、心理学において用いられるデータ分析の手続きとして分散分析と並んで頻繁に用いられるものは回帰分析であり、近年では回帰分析をより洗練した統計手法（階層的因子分析や媒介分析, SEM）が変数間の関係を検討する研究において用いられる (Blanca et al., 2018)。これらの手法は全て変数間の相関係数を基にして分析が進められる。したがって、現状の心理学研究において相関係数を用いる手法の多様性を鑑みると、相関係数に生じたバイアスへの適切な対処法を用いることの意義は大きいだろう。

研究 2 が示すように、変数が 2 変量正規分布に従うことが想定される場合には polychoric 相関係数の利用が推奨されるだろう。離散化の際のカテゴリ数が多い場合には Pearson の相関係数を用いても問題ないとする主張もあるが (e.g., 萩生田・繁柵, 1996)、研究 1 で示すように離散化の際の閾値が多様な値をとりうる場合には相関係数により大きなバイアスが生じることから、polychoric 相関係数の利用が適

さない場合<sup>1</sup>をのぞいて順序カテゴリ変数から相関係数を算出する際には polychoric 相関係数を用いた方が無難であろう。

相関係数のバイアスの帰結として本研究では信頼性係数への影響を評価した。研究 3 において Cronbach の  $\alpha$  は、床・天井効果を示す項目数が多いほど大きく低下することを示したが、これは離散化の際に整った閾値を用いるより多様な閾値を用いた場合に相関係数により大きなバイアスが生じることを示した研究 1 と整合的である。(5.7) が示すように  $\alpha$  は相関係数の関数だからである。また、研究 4 の結果からは、Yang & Green (2015) や Kim et al. (2020) などの先行研究が示すように順序カテゴリデータに対する因子分析に基づく信頼性係数である非線形 SEM 係数が小さいバイアスで信頼性係数を推定できることが示された。項目のカテゴリ数が少ない場合には有用な選択肢として用いることが可能であろう。その一方で研究 4-a は、モデルが誤って特定された場合には非線形 SEM 係数に大きなバイアスが生じうることも示唆しており、この係数の可能性とともに留意点が論じられた。これは非線形 SEM 係数に限った話ではないが、モデルに基づく信頼性係数を利用する場合には、モデルの注意深い検討が肝要であること (Savalei & Reise, 2019) を再確認した形である。

研究 2-a や研究 4-a の新奇性はバイアスへの対処法そのものではなく、バイアスへの対処法が上手く機能する条件を指摘したことにある。研究 2-a の結果からは「polychoric 相関係数はカテゴリ化による相関係数のバイアスを補正することができる（ただし、カテゴリ数が少なくかつ 2 変数が逆方向に歪んでいる場合を除いて）」と言え、研究 4-a の結果からは「非線形 SEM 係数は順序カテゴリデータにおける信頼性を上手く推定することができる（ただしモデルが正しく特定されている限りにおいて）」と言える。心理学の研究ではしばしば提案された手法が上手く機能するための条件が不明なまま応用研究で頻繁に用いられることがあるが、手法の限界を正しく指摘すること、すなわち上の文言のかっこ内の但し書きの部分を明らかにする

---

<sup>1</sup> 離散化前の分布が 2 変量正規分布に従っていない場合や潜在変数がそもそも離散的な場合が挙げられる。後者について、例えば Jansen & van der Maas (1997) は Siegler の天秤課題の回答パターンを説明するために、子どもが課題解決に用いているルールを離散的な潜在クラスとしてモデル化している。

ことは、適切なデータ解析のために地味ではあるが必須の作業であろう。より具体的には、方法論の研究者には提案された手法が現実のデータ解析の多様な条件下で適用可能かを検討していくことが求められるだろうし、提案された手法を実際のデータ解析に用いる研究者はそうした検討の結果を参照して、手法が適用可能な条件にあるかを確認する慎重さが求められるだろう。

本研究の結果から、Likert 型の項目<sup>2</sup>を用いて相関係数および信頼性係数を計算する際の留意点は以下の3点に集約できる。第1に、カテゴリ数が多い場合でも相関係数に大きなバイアスが生じる場合があるため順序カテゴリデータから相関係数を算出する場合には、基本的には polychoric 相関係数の利用が推奨される。その際には2変量正規分布の仮定に留意し、仮定が満たされないと想定される場合には利用は控えることが望ましい。第2に、尺度作成場面で信頼性係数の  $\alpha$  を用いて項目数の検討を行う場合には、尺度内の項目に床・天井効果を示す項目が多いと  $\alpha$  は大きく低下するので、これらの存在の有無や項目に占める割合を確認した上で項目数の設計を行うことである。第3に、信頼性の尺度として非線形 SEM 係数を用いる際には係数の算出の基にしたモデルの適切性を複数の異なる適合度指標、既存の理論との整合性などの観点から総合的に検討した上で用いることである。

### 6.3 本研究の課題と今後の展望

本研究では、順序カテゴリデータから計算される相関係数のバイアスおよびその対処法の評価を行った。しかし、以下に挙げるような課題が残されている。第1に、相関係数を用いた分析には様々なものがあるが、それら全てを扱うことができていない点である。相関係数を用いる研究において利用される頻度の高い回帰分析や、尺度作成の際に多用される因子分析において、カテゴリ化によるバイアスがどのよう

---

<sup>2</sup> 第1章において区別したように Likert 型の項目と Likert 尺度は異なるものであり、本研究が対象としたバイアスは前者から計算される相関係数に生じるものである。心理学の研究では複数の項目の得点を合計（あるいは平均）した尺度得点を用い、尺度得点間の相関係数を利用するものも多い。これらの尺度得点は厳密に言えば順序カテゴリデータだが、Carifio & Perla (2008) が指摘するように Likert 尺度の得点を間隔尺度と見做しても問題ないケースは多く、Likert 尺度の得点間の相関係数は Likert 型の項目から計算される相関係数よりもロバストであると考えられている。もっとも、各尺度の項目数やカテゴリ化の際の閾値によっては大きなバイアスが生じる可能性もあるので、これは別途詳細な検討が必要な課題であろう。



な影響を及ぼすかについて検討することが今後必要であろう。

第 2 に、本研究では 2 変量正規分布の仮定が満たされた場合には polychoric 相関係数が現状だと対処法の第 1 選択肢であると論じたが、polychoric 相関係数の算出には Olsson (1979a) による方法においては 1 ステップ法と 2 ステップ法があり、EM アルゴリズムを用いて推定するもの (Shiina et al., 2018)、ベイズ推定を用いるもの (Choi, Kim, Chen, & Dannels, 2011)、など様々なヴァリエーションが存在している。これらの推定方法によって polychoric 相関係数のパフォーマンスがどう異なるのか、この違いがその後の分析においてどう影響するのか検討が今後必要になるだろう。

## 参考文献

- Adler, D., Glaser, C., Nenadic, O., Oehlschlagel, J., & Zucchini, W. (2018). ff: Memory-efficient storage of large data on disk and fast access functions (version 2.2-14). <https://CRAN.R-project.org/package=ff>
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137–143. <https://doi.org/10.1007/s11336-008-9100-1>
- Blanca, M. J., Alarcón, R., & Bono, R. (2018). Current practices in data analysis procedures in psychology: What has changed? *Frontiers in Psychology*, *9*, 2558. <https://doi.org/10.3389/fpsyg.2018.02558>
- Bollen, K., & Barb, K. (1981). Pearson's R and coarsely categorized measures. *American Sociological Review*, *46*(2), 232–239. <https://doi.org/10.2307/2094981>
- Brockett, P. L., & Levine, A. (1977). On a characterization of ridits. *The Annals of Statistics*, *5*(6), 1245–1248. <https://doi.org/10.1214/aos/1176344010>
- Bross, I. D. J. (1958). How to use ridity analysis. *Biometrics*, *14*(1), 18–38. <https://doi.org/10.2307/2527727>
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150–1152. <https://doi.org/10.1111/j.1365-2923.2008.03172.x>
- Chalmers, R. P. (2018). On misconceptions and the limited usefulness of ordinal alpha. *Educational and Psychological Measurement*, *78*(6), 1056–1071. <https://doi.org/10.1177/0013164418791111>

- [//doi.org/10.1177/0013164417727036](https://doi.org/10.1177/0013164417727036)
- Choi, J., Kim, S., Chen, J., & Dannels, S. (2011). A comparison of maximum likelihood and bayesian estimation for polychoric correlation using Monte Carlo simulation. *Journal of Educational and Behavioral Statistics*, *36*(4), 523–549. <https://doi.org/10.3102/1076998610381398>
- Clason, D., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, *61*(4), 31–35. <https://doi.org/10.5032/JAE.1994.04031>
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. <https://doi.org/10.1177/014662168300700301>
- Coombs, C. H., Raiffa, H., & Thrall, R. M. (1954). Some views on mathematical models and measurement theory. *Psychological Review*, *61*(2), 132–144. <https://doi.org/10.1037/h0063044>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. <https://doi.org/10.1007/BF02310555>
- David, H. A., & Nagaraja, H. N. (2003). *Order Statistics* (3rd ed.). Hoboken, NJ: Wiley-Interscience.
- Duursma, R. (2019). lgrdata: Example datasets for a learning guide to R (version 0.1.1). <https://CRAN.R-project.org/package=lgrdata>
- Edmondson, D. (2005). Likert scales: A history. *Proceedings of the Conference on Historical Analysis and Research in Marketing*, *12*, 127–133.
- Ellson, D. G., & Ellson, E. C. (1953). Historical note on the rating scale. *Psychological Bulletin*, *50*(5), 383–384. <https://doi.org/10.1037/h0054149>
- Everitt, P. F. (1910). Tables of the tetrachoric functions for fourfold correlation talbes. *Biometrika*, *7*(4), 437–451. <https://doi.org/10.1093/biomet/7.4.>

- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. <https://doi.org/10.1080/00273170701382864>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling.. In *Structural Equation Modeling: A Second Course* (2nd ed.)(pp. 439–492). Charlotte, NC: Information Age Publishing.
- Fox, J. (2019). polycor: Polychoric and polyserial correlations (version 0.7-10). <https://CRAN.R-project.org/package=polycor>
- Furr, R. M. (2018). *Psychometrics: An Introduction* (3rd ed.). Los Angeles, CA: SAGE Publications, Inc.
- Gadermann, A., Guhn, M., & Zumbo, B. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1–13. <https://doi.org/10.7275/n560-j767>
- Galton, F. (1883). *Inquiries into Human Faculty and Its Development*. London: Macmillan.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2018). mvtnorm: Multivariate normal and t distributions (version 1.0-8). <https://CRAN.R-project.org/package=mvtnorm>
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2020). mvtnorm: Multivariate normal and t distributions (version 1.1-0). <https://CRAN.R-project.org/package=mvtnorm>
- Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*(1), 155–167. <https://doi.org/10.1007/s11336-008-9099-3>
- Grønneberg, S., & Foldnes, N. (2019). A problem with discretizing Vale–Maurelli in simulation studies. *Psychometrika*, *84*(2), 554–561. <https://doi.org/10.>

1007/s11336-019-09663-8

- Guilford, J. P. (1954). *Psychometric Methods* (2nd ed.). New York, NY: McGraw-Hill. (ギルホード, J. P. 秋重 義治 (訳)(1959). 精神測定法 培風館)
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. <https://doi.org/10.1007/BF02288892>
- 萩生田 伸子・繁梲 算男 (1996). 順序付きカテゴリカルデータへの因子分析の適用に関するいくつかの注意点 心理学研究, *67*(1), 1–8. <https://doi.org/10.4992/jjpsy.67.1>
- Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007). Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *The Lancet Neurology*, *6*(12), 1094–1105. [https://doi.org/10.1016/S1474-4422\(07\)70290-9](https://doi.org/10.1016/S1474-4422(07)70290-9)
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, *60*(4), 523–531. <https://doi.org/10.1177/00131640021970691>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, structural equation modeling. *Structural Equation Modeling*, *6*(1), 1–55. <https://doi.org/http://dx.doi.org/10.1080/10705519909540118>
- Hunt, T. (2013). Lambda4: Collection of internal consistency reliability coefficients (version 3.0). <https://CRAN.R-project.org/package=Lambda4>
- 岩坪 秀一 (1987). 数量化法の基礎 朝倉書店
- Jackson, D. L., Gillaspay, J. A. J., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations.

- Psychological Methods*, 14(1), 6–23. <https://doi.org/10.1037/a0014694>
- Jansen, B. R. J., & van der Maas, H. L. J. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review*, 17(3), 321–357. <https://doi.org/10.1006/drev.1997.0437>
- 狩野 裕 (2002). 再討論——誤差共分散の利用と特殊因子の役割—— 行動計量学, 29(2), 182–197. <https://doi.org/10.2333/jbhmk.29.182>
- Kano, Y., & Azuma, Y. (2003). Use of SEM programs to precisely measure scale reliability. In Yanai, H., Okada, A., Shigemasu, K., Kano, Y., & Meulman, J. J.(eds.) *New Developments in Psychometrics*(pp. 141–148). Tokyo: Springer. [https://doi.org/10.1007/978-4-431-66996-8\\_14](https://doi.org/10.1007/978-4-431-66996-8_14)
- Kendall, M. G., & Stuart, A. (1979). *The Advanced Theory of Statistics, Volume 2: Inference and Relationship* (4th ed.). London: Charles Griffin & Company.
- Kim, S., Lu, Z., & Cohen, A. S. (2020). Reliability for tests with items having different numbers of ordered categories. *Applied Psychological Measurement*, 44(2), 137–149. <https://doi.org/10.1177/0146621619835498>
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160. <https://doi.org/10.1007/BF02288391>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1–55.
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Mass: Addison-Wesley Pub. Co.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. <https://doi.org/10.1037/1082-989x.7.1.19>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler’s (1999) findings.

- Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341.  
[https://doi.org/10.1207/s15328007sem1103\\_2](https://doi.org/10.1207/s15328007sem1103_2)
- Martin, W. S. (1973). The effects of scaling on the correlation coefficient: A test of validity. *Journal of Marketing Research*, 10(3), 316–318. <https://doi.org/10.2307/3149702>
- Martin, W. S. (1978). Effects of scaling on the correlation coefficient: Additional considerations. *Journal of Marketing Research*, 15(2), 304–308. <https://doi.org/10.2307/3151268>
- Masters, E. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement*, 11(1), 49–53. <https://doi.org/10.1111/j.1745-3984.1974.tb00970.x>
- McDonald, R. P. (1985). *Factor Analysis and Related Methods*. New York, NY: Psychology Press.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- McReynolds, P., & Ludwig, K. (1987). On the history of rating scales. *Personality and Individual Differences*, 8(2), 281–283. [https://doi.org/10.1016/0191-8869\(87\)90188-7](https://doi.org/10.1016/0191-8869(87)90188-7)
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, 100(3), 398–407. <https://doi.org/10.1037/0033-2909.100.3.398>
- Morrison, D. G. (1972). Regressions with discrete dependent variables: The effect on  $R^2$ . *Journal of Marketing Research*, 9(3), 338–340. <https://doi.org/10.2307/3149551>
- 村瀬 雄平 (1917). 智能の遺傳——日本人についての研究—— 心理学研究会出版部
- Muthén, B. (1984). A general structural equation model with dichotomous, or-

- dered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132. <https://doi.org/10.1007/BF02294210>
- 西里 静彦 (1982). 質的データの数量化——双対尺度法とその応用—— 朝倉書店
- Nishisato, S., & Torii, Y. (1971). Effects of categorizing continuous normal variables on product-moment correlation. *Japanese Psychological Research*, 13(1), 45–49. <https://doi.org/10.4992/psycholres1954.13.45>
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
- O’Brien, R. M. (1981). Reducing grouping distortion in rank category variables. *American Journal of Political Science*, 25(3), 605–616. <https://doi.org/10.2307/2110821>
- O’Brien, R. M. (1983). Rank order versus rank category measures of continuous variables. *American Sociological Review*, 48(2), 284–286. <https://doi.org/10.2307/2095113>
- O’Brien, R. M. (1985). The relationship between ordinal measures and their underlying values: Why all the disagreement? *Quality and Quantity*, 19(3), 265–277. <https://doi.org/10.1007/BF00170998>
- 岡田 謙介 (2011). クロンバックの  $\alpha$  に代わる信頼性の推定法について——構造方程式モデリングによる方法・McDonald の  $\omega$  の比較—— 日本テスト学会誌, 7(1), 38–50.
- 岡田 謙介 (2015). 心理学と心理測定における信頼性について——Cronbach の  $\alpha$  係数とは何なのか, 何でないのか—— 教育心理学年報, 54, 71–83. <https://doi.org/10.5926/arepj.54.71>
- 岡本 浩一 (1991). 社会心理学における測定 市川 伸一 (編) 心理測定法への招待——測定からみた心理学入門——(pp. 215–235) サイエンス社
- Olsson, U. (1979a). Maximum likelihood estimation of the polychoric correlation. *Psychometrika*, 44(4), 443–460. <https://doi.org/10.1007/BF02296207>



- Olsson, U. (1979b). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research*, *14*(4), 485–500. [https://doi.org/10.1207/s15327906mbr1404\\_7](https://doi.org/10.1207/s15327906mbr1404_7)
- 小野島 昂洋・椎名 乾平 (2019). カテゴリカルな教育データと相関係数についての留意点 コンピュータ&エデュケーション, *47*, 25–30. <https://doi.org/10.14949/konpyutariyoukyouiku.47.25>
- 小野島 昂洋・椎名 乾平 (2020). 変数のカテゴリ化によって生じる相関係数のバイアスがクロンバックの  $\alpha$  へと及ぼす影響 日本教育心理学会第 62 回総会発表論文集, *62*, 371.
- 小野島 昂洋・椎名 乾平 (2021). 順序カテゴリデータへの確認的因子分析に基づく信頼性係数の評価——モデルが正しく特定された場合と誤特定された場合の比較—— 教育心理学研究, *69*(3), 281–296.
- Onoshima, T., Shiina, K., Ueda, T., & Kubo, S. (2019). Decline of Pearson's  $r$  with categorization of variables: A large-scale simulation. *Behaviormetrika*, *46*(2), 380–399. <https://doi.org/10.1007/s41237-019-00089-1>
- 大内 善広・上田 卓司・椎名 乾平・岡田 いずみ (2012). カテゴリ数異なる順序カテゴリ尺度同士の相関係数の性質 学術研究 (人文科学・社会科学編), *60*, 93–103.
- Pearson, K. (1896). Mathematical contributions to the theory of evolution, III: Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *187*, 253–318.
- Pearson, K. (1900). Mathematical contributions to the theory of evolution, VII: On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *195*, 1–47. <https://doi.org/10.1098/rsta.1900.0022>
- Pearson, K. (1913). On the measurement of the influence of “broad categories”

- on correlation. *Biometrika*, *9*, 116–139. <https://doi.org/10.2307/2331803>
- Pearson, K., & Pearson, E. S. (1922). On polychoric coefficients of correlation. *Biometrika*, *14*, 127–156. <https://doi.org/10.2307/2331858>
- Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical Procedures and Their Mathematical Bases*. New York, NY: McGraw-Hill.
- R Core Team (2018). R: A language and environment for statistical computing (version 3.5.0). <https://www.R-project.org/>
- R Core Team (2020). R: A language and environment for statistical computing (version 3.6.3). <https://www.R-project.org/>
- Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, *79*(1), 200–210. <https://doi.org/10.1177/0013164417725127>
- Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 195–212. [https://doi.org/10.1207/S15328007SEM0902\\_3](https://doi.org/10.1207/S15328007SEM0902_3)
- Revelle, W., & Condon, D. M. (2019). Reliability from  $\alpha$  to  $\omega$ : A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. <https://doi.org/10.1037/pas0000754>
- Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2013). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research*, *43*(1), 73–97. <https://doi.org/10.1177/0049124113509605>
- Ritchie-Scott, A. (1918). The correlation coefficient of a polychoric table. *Biometrika*, *12*, 93–133. <https://doi.org/10.2307/2331931>
- Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. New Jersey: Princeton University Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling (version 0.6-5). *Journal of Statistical Software*, *48*(2), 1–36. <https://doi.org/>

10.18637/jss.v048.i02

- Rosseel, Y. (2014). Structural equation modeling with categorical variables. [https://personality-project.org/r/tutorials/summerschool.14/rosseel\\_sem\\_cat.pdf](https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_cat.pdf)
- Rucci, A. J., & Tweney, R. D. (1980). Analysis of variance and the “second discipline” of scientific psychology: A historical account. *Psychological Bulletin*, 87(1), 166–184. <https://doi.org/10.1037/0033-2909.87.1.166>
- Savalei, V., & Reise, S. P. (2019). Don’t forget the model in your model-based reliability coefficients: A reply to McNeish (2018). *Collabra: Psychology*, 5(1), 36. <https://doi.org/10.1525/collabra.247>
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53–61. <https://doi.org/10.1007/BF02289209>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- 椎名 乾平 (2016). 相関係数の起源と多様な解釈 心理学評論, 59(4), 415–444. [https://doi.org/10.24602/sjpr.59.4\\_415](https://doi.org/10.24602/sjpr.59.4_415)
- 椎名 乾平 (2017). 知られざる Pearson (1913) の Broad-Category 補正式——カテゴリー尺度による相関係数の問題—— 学術研究 (人文科学・社会科学編), 65, 121–142.
- Shiina, K., Ueda, T., & Kubo, S. (2018). Polychoric correlations for ordered categories using the EM algorithm. In Wiberg, M., Culpepper, S. A., Janssen, R., González, J., & Molenaar, D.(eds.) *Quantitative Psychology*(pp. 247–259). New York, NY: Springer International Publishing. <https://doi.org/10.1007/978-3-319-56294-0>
- 清水 和秋 (2007).  $\alpha$  はやめて  $\omega$  にしよう 日本心理学会大会発表論文集, 71, 2PM049. [https://doi.org/10.4992/pacjpa.71.0\\_2PM049](https://doi.org/10.4992/pacjpa.71.0_2PM049)
- 清水 和秋・吉田 昂平 (2008). Rosenberg 自尊感情尺度のモデル化——Wording と項目配置の影響の検討—— 関西大学社会学部紀要, 39(2), 69–97.

- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>
- Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for “optimal” cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education*, *74*(3), 267–288. <https://doi.org/10.3200/JEXE.74.3.267-288>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. <https://doi.org/10.2307/1412159>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *1904-1920*, *3*(3), 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>
- 高本 真寛・服部 環 (2015). 国内の心理尺度作成論文における信頼性係数の利用動向 心理学評論, *58*(2), 220-235. [https://doi.org/10.24602/sjpr.58.2\\_220](https://doi.org/10.24602/sjpr.58.2_220)
- 田中 良久 (1977). 心理学的測定法, 第2版 東京大学出版会
- Thorndike, E. L. (1904). *An Introduction to the Theory of Mental and Social Measurements*. New York: Science Press.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*(4), 529–554. <https://doi.org/10.1086/214483>
- Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York: Wiley.
- Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice*, *16*(4), 8–14. <https://doi.org/10.1111/j.1745-3992.1997.tb00603.x>
- Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling: A Multidisciplinary Jour-*

*nal*, 17(1), 66–81. <https://doi.org/10.1080/10705510903438963>

Yang, Y., & Green, S. B. (2015). Evaluation of structural equation modeling estimates of reliability for scales with ordered categorical items. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 11(1), 23–34. <https://doi.org/10.1027/1614-2241/a000087>

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21–29. <https://doi.org/10.22237/jmasm/1177992180>

付録 A

## 研究 1 の補足資料

Table A.1  
 一様分布条件の平均相関係数 ( $N = 64$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.96	0.98	1
m2n2	.000	.056	.113	.170	.228	.288	.348	.412	.476	.545	.590	.600	.619
m2n3	.000	.063	.126	.190	.256	.322	.388	.459	.533	.611	.661	.674	.694
m2n4	.000	.065	.131	.198	.265	.334	.406	.477	.555	.634	.683	.704	.720
m2n5	.000	.066	.134	.202	.270	.340	.412	.487	.564	.649	.698	.717	.736
m2n6	.000	.067	.136	.204	.274	.345	.418	.493	.572	.654	.708	.724	.746
m2n7	.000	.068	.137	.206	.276	.348	.421	.498	.577	.661	.713	.731	.753
m3n3	.000	.070	.142	.213	.285	.360	.436	.515	.597	.684	.739	.760	.780
m3n4	.000	.073	.147	.221	.297	.375	.454	.536	.622	.711	.768	.788	.809
m3n5	.000	.074	.150	.226	.303	.382	.463	.547	.634	.726	.783	.805	.826
m3n6	.000	.075	.152	.229	.306	.386	.468	.553	.642	.733	.795	.816	.834
m3n7	.000	.076	.153	.231	.309	.390	.472	.557	.648	.741	.801	.823	.844
m4n4	.000	.076	.153	.230	.309	.389	.471	.556	.647	.739	.800	.820	.841
m4n5	.000	.077	.156	.235	.315	.397	.481	.568	.659	.755	.817	.837	.859
m4n6	.000	.078	.158	.238	.319	.401	.487	.575	.667	.764	.825	.847	.869
m4n7	.000	.079	.159	.240	.322	.405	.491	.580	.673	.772	.834	.854	.876
m5n5	.000	.079	.159	.240	.321	.405	.491	.580	.672	.770	.834	.855	.878
m5n6	.000	.080	.161	.242	.325	.410	.497	.587	.681	.780	.843	.864	.888
m5n7	.000	.081	.162	.244	.328	.413	.501	.591	.686	.786	.851	.873	.895
m6n6	.000	.081	.163	.245	.329	.415	.503	.594	.689	.789	.855	.877	.899
m6n7	.000	.082	.164	.247	.332	.418	.507	.599	.695	.796	.861	.883	.906
m7n7	.000	.082	.166	.249	.334	.422	.511	.604	.701	.803	.869	.891	.914

Table A.2

一様分布条件の平均相関係数 ( $N = 256$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.96	0.98	1
m2n2	.000	.056	.113	.170	.229	.288	.348	.411	.477	.544	.591	.602	.620
m2n3	.000	.063	.127	.190	.256	.322	.391	.462	.534	.612	.657	.677	.689
m2n4	.000	.065	.132	.198	.266	.336	.406	.480	.557	.635	.685	.705	.722
m2n5	.000	.067	.134	.202	.272	.342	.414	.490	.566	.649	.700	.718	.735
m2n6	.000	.067	.136	.205	.274	.346	.419	.495	.573	.656	.710	.728	.746
m2n7	.000	.068	.137	.206	.277	.348	.423	.498	.579	.662	.716	.735	.753
m3n3	.000	.070	.142	.214	.287	.361	.437	.517	.599	.685	.740	.759	.780
m3n4	.000	.073	.148	.222	.298	.375	.454	.537	.621	.711	.770	.789	.809
m3n5	.000	.075	.151	.226	.304	.382	.463	.548	.634	.728	.785	.805	.826
m3n6	.000	.075	.152	.229	.308	.387	.469	.554	.642	.736	.796	.815	.835
m3n7	.000	.076	.153	.231	.310	.390	.473	.559	.649	.742	.801	.822	.843
m4n4	.000	.076	.153	.230	.309	.390	.472	.557	.647	.739	.799	.822	.843
m4n5	.000	.078	.156	.235	.316	.398	.482	.570	.660	.756	.815	.836	.859
m4n6	.000	.078	.158	.238	.320	.402	.488	.576	.668	.765	.825	.849	.870
m4n7	.000	.079	.160	.240	.322	.406	.492	.581	.673	.772	.833	.855	.877
m5n5	.000	.079	.159	.240	.322	.406	.492	.581	.673	.771	.834	.854	.878
m5n6	.000	.080	.162	.243	.326	.410	.498	.587	.682	.781	.844	.865	.888
m5n7	.000	.081	.163	.245	.329	.414	.502	.592	.687	.788	.850	.873	.894
m6n6	.000	.081	.163	.246	.330	.416	.504	.595	.690	.791	.855	.877	.900
m6n7	.000	.082	.165	.248	.333	.419	.508	.600	.696	.797	.862	.883	.907
m7n7	.000	.082	.166	.250	.335	.423	.512	.605	.702	.803	.869	.891	.913



Table A.3  
 一様分布条件の平均相関係数 ( $N = 1024$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.96	0.98	1
m2n2	.000	.056	.113	.170	.228	.288	.349	.411	.477	.547	.591	.605	.618
m2n3	.000	.063	.127	.191	.256	.322	.391	.461	.533	.612	.659	.679	.694
m2n4	.000	.065	.132	.198	.266	.335	.406	.481	.555	.636	.685	.704	.724
m2n5	.000	.067	.134	.202	.271	.342	.414	.488	.564	.648	.700	.720	.737
m2n6	.000	.067	.136	.205	.275	.347	.419	.495	.575	.655	.711	.726	.746
m2n7	.000	.068	.137	.207	.277	.349	.423	.500	.580	.661	.714	.736	.751
m3n3	.000	.070	.142	.214	.287	.362	.438	.517	.599	.686	.742	.760	.778
m3n4	.000	.073	.148	.222	.298	.375	.455	.537	.622	.712	.769	.788	.811
m3n5	.000	.075	.150	.227	.304	.383	.464	.548	.635	.726	.786	.805	.827
m3n6	.000	.076	.152	.229	.308	.387	.470	.555	.643	.736	.795	.815	.838
m3n7	.000	.076	.154	.231	.310	.391	.474	.559	.649	.742	.802	.822	.844
m4n4	.000	.076	.153	.231	.309	.390	.472	.557	.647	.741	.800	.821	.842
m4n5	.000	.078	.156	.236	.316	.398	.482	.569	.660	.755	.814	.838	.859
m4n6	.000	.078	.158	.238	.320	.403	.488	.576	.668	.765	.827	.848	.870
m4n7	.000	.079	.160	.241	.322	.406	.492	.581	.674	.772	.834	.855	.878
m5n5	.000	.079	.159	.240	.322	.405	.492	.580	.674	.771	.833	.857	.877
m5n6	.000	.080	.161	.243	.326	.411	.498	.588	.682	.781	.843	.865	.888
m5n7	.000	.081	.163	.245	.329	.414	.502	.593	.688	.787	.851	.872	.895
m6n6	.000	.081	.163	.246	.330	.416	.504	.595	.690	.791	.854	.877	.899
m6n7	.000	.082	.165	.248	.333	.419	.508	.600	.696	.797	.862	.884	.907
m7n7	.000	.082	.166	.250	.335	.423	.513	.605	.702	.804	.868	.891	.913

Table A.4  
等距離条件の平均相関係数 ( $N = 64$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.96	0.98	1
m2n2	.000	.062	.126	.191	.257	.327	.401	.481	.572	.680	.762	.793	.827
m2n3	.000	.069	.140	.211	.284	.361	.442	.528	.625	.736	.812	.839	.865
m2n4	.000	.071	.144	.218	.293	.372	.455	.544	.641	.752	.827	.853	.880
m2n5	.000	.072	.146	.221	.298	.378	.461	.551	.649	.759	.833	.858	.883
m2n6	.000	.073	.148	.223	.301	.381	.465	.555	.653	.763	.835	.860	.885
m2n7	.000	.074	.149	.225	.302	.383	.467	.557	.656	.765	.836	.861	.885
m3n3	.000	.076	.154	.232	.312	.395	.482	.575	.675	.788	.865	.893	.924
m3n4	.000	.079	.159	.239	.322	.407	.497	.592	.694	.808	.885	.912	.941
m3n5	.000	.080	.161	.243	.327	.414	.504	.600	.703	.817	.893	.921	.948
m3n6	.000	.080	.162	.245	.330	.417	.508	.604	.708	.822	.898	.924	.952
m3n7	.000	.081	.163	.247	.332	.419	.511	.607	.711	.825	.900	.926	.953
m4n4	.000	.081	.163	.246	.331	.419	.511	.607	.711	.825	.901	.928	.957
m4n5	.000	.082	.166	.250	.337	.425	.518	.616	.720	.835	.911	.938	.966
m4n6	.000	.083	.167	.253	.340	.429	.522	.621	.726	.840	.915	.942	.970
m4n7	.000	.083	.168	.254	.341	.431	.525	.624	.729	.843	.918	.945	.973
m5n5	.000	.083	.168	.254	.341	.431	.525	.623	.728	.843	.918	.944	.973
m5n6	.000	.084	.170	.256	.344	.435	.529	.628	.733	.848	.923	.950	.978
m5n7	.000	.085	.171	.257	.346	.437	.532	.631	.737	.851	.926	.953	.980
m6n6	.000	.085	.171	.258	.347	.438	.533	.632	.738	.852	.927	.953	.981
m6n7	.000	.085	.172	.259	.349	.440	.535	.635	.741	.856	.930	.956	.984
m7n7	.000	.086	.173	.261	.350	.442	.538	.638	.744	.858	.932	.959	.986

Table A.5  
等距離条件の平均相関係数 ( $N = 256$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.96	0.98	1
m2n2	.000	.062	.126	.191	.258	.327	.401	.481	.572	.681	.761	.795	.825
m2n3	.000	.069	.140	.211	.285	.361	.442	.529	.625	.736	.813	.840	.866
m2n4	.000	.071	.144	.218	.294	.372	.455	.544	.642	.753	.828	.854	.880
m2n5	.000	.073	.147	.221	.298	.378	.462	.551	.649	.760	.833	.858	.883
m2n6	.000	.073	.148	.223	.301	.381	.466	.555	.654	.763	.835	.860	.885
m2n7	.000	.074	.149	.225	.303	.383	.468	.558	.656	.765	.837	.861	.885
m3n3	.000	.076	.154	.232	.312	.395	.483	.575	.676	.789	.865	.894	.924
m3n4	.000	.079	.159	.240	.322	.408	.497	.592	.695	.809	.885	.913	.941
m3n5	.000	.080	.161	.243	.327	.414	.505	.600	.704	.818	.894	.921	.949
m3n6	.000	.081	.163	.246	.330	.418	.509	.605	.709	.823	.898	.925	.952
m3n7	.000	.081	.164	.247	.332	.420	.511	.608	.712	.825	.900	.927	.953
m4n4	.000	.081	.164	.247	.332	.420	.511	.608	.712	.826	.901	.929	.958
m4n5	.000	.082	.166	.251	.337	.426	.519	.617	.721	.836	.911	.938	.966
m4n6	.000	.083	.168	.253	.340	.430	.523	.621	.726	.841	.916	.943	.970
m4n7	.000	.084	.169	.255	.342	.432	.526	.624	.730	.844	.919	.945	.973
m5n5	.000	.084	.169	.254	.342	.432	.526	.624	.729	.843	.918	.945	.973
m5n6	.000	.084	.170	.257	.345	.436	.530	.629	.734	.849	.923	.950	.978
m5n7	.000	.085	.171	.258	.347	.438	.533	.632	.738	.852	.927	.953	.980
m6n6	.000	.085	.171	.259	.347	.439	.534	.633	.739	.853	.927	.953	.981
m6n7	.000	.086	.173	.260	.349	.441	.536	.636	.742	.856	.931	.957	.984
m7n7	.000	.086	.173	.261	.351	.443	.539	.639	.745	.859	.933	.959	.986

Table A.6  
等距離条件の平均相関係数 ( $N = 1024$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.96	0.98	1
m2n2	.000	.062	.126	.191	.258	.327	.401	.481	.572	.681	.762	.794	.828
m2n3	.000	.069	.140	.211	.285	.361	.442	.529	.625	.736	.813	.840	.867
m2n4	.000	.071	.144	.218	.294	.372	.455	.544	.642	.753	.828	.854	.880
m2n5	.000	.073	.147	.222	.298	.378	.462	.551	.649	.760	.833	.858	.883
m2n6	.000	.073	.148	.224	.301	.381	.466	.556	.654	.763	.835	.860	.885
m2n7	.000	.074	.149	.225	.303	.383	.468	.558	.656	.765	.837	.861	.885
m3n3	.000	.076	.154	.232	.313	.396	.483	.575	.676	.789	.866	.894	.925
m3n4	.000	.079	.159	.240	.323	.408	.498	.592	.695	.809	.885	.913	.941
m3n5	.000	.080	.161	.243	.327	.414	.505	.601	.704	.818	.894	.921	.949
m3n6	.000	.081	.163	.246	.330	.418	.509	.605	.709	.823	.898	.925	.952
m3n7	.000	.081	.164	.247	.332	.420	.512	.608	.712	.826	.900	.927	.954
m4n4	.000	.081	.164	.247	.332	.420	.512	.608	.712	.826	.901	.929	.958
m4n5	.000	.082	.166	.251	.337	.426	.519	.617	.721	.836	.911	.938	.966
m4n6	.000	.083	.168	.253	.340	.430	.523	.622	.727	.841	.916	.943	.970
m4n7	.000	.084	.169	.255	.342	.432	.526	.625	.730	.844	.919	.945	.973
m5n5	.000	.084	.169	.254	.342	.432	.526	.624	.729	.843	.918	.945	.973
m5n6	.000	.084	.170	.257	.345	.436	.530	.629	.735	.849	.924	.950	.978
m5n7	.000	.085	.171	.258	.347	.438	.533	.632	.738	.852	.927	.953	.980
m6n6	.000	.085	.172	.259	.348	.439	.534	.633	.739	.853	.927	.954	.981
m6n7	.000	.086	.173	.260	.350	.441	.537	.637	.742	.856	.931	.957	.984
m7n7	.000	.086	.173	.261	.351	.444	.539	.639	.745	.859	.933	.959	.986

Table A.7

一様分布条件の相関係数の標準偏差 ( $N = 64$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98	1	
m2n2	.126	.126	.125	.123	.122	.120	.120	.124	.137	.163	.192	.206	.224
m2n3	.126	.126	.125	.123	.120	.118	.115	.115	.119	.135	.154	.164	.179
m2n4	.126	.125	.124	.122	.119	.116	.112	.110	.112	.124	.139	.146	.157
m2n5	.126	.125	.124	.122	.119	.115	.111	.108	.108	.116	.130	.136	.143
m2n6	.126	.125	.124	.122	.118	.114	.110	.106	.105	.112	.123	.128	.135
m2n7	.126	.125	.124	.122	.118	.114	.109	.105	.104	.109	.118	.125	.130
m3n3	.126	.126	.124	.122	.119	.115	.111	.107	.106	.112	.125	.131	.138
m3n4	.126	.125	.124	.121	.118	.113	.108	.102	.098	.102	.109	.115	.123
m3n5	.126	.125	.124	.121	.117	.112	.106	.100	.095	.095	.102	.105	.110
m3n6	.126	.125	.124	.121	.117	.112	.105	.098	.092	.093	.096	.098	.107
m3n7	.126	.125	.124	.121	.117	.111	.105	.097	.091	.088	.092	.095	.099
m4n4	.126	.125	.124	.121	.117	.111	.104	.097	.090	.091	.095	.099	.105
m4n5	.126	.125	.124	.120	.116	.110	.103	.095	.087	.083	.086	.090	.095
m4n6	.126	.125	.124	.120	.116	.110	.102	.093	.084	.079	.081	.084	.088
m4n7	.126	.125	.123	.120	.115	.109	.101	.092	.082	.075	.075	.080	.084
m5n5	.126	.125	.123	.120	.115	.109	.101	.092	.083	.077	.077	.079	.082
m5n6	.126	.125	.123	.120	.115	.108	.100	.090	.079	.072	.072	.074	.076
m5n7	.126	.125	.123	.120	.115	.108	.099	.089	.078	.069	.067	.068	.072
m6n6	.126	.125	.123	.120	.114	.108	.099	.088	.077	.068	.065	.067	.068
m6n7	.126	.125	.123	.119	.114	.107	.098	.087	.075	.064	.062	.062	.065
m7n7	.126	.125	.123	.119	.114	.107	.097	.086	.073	.061	.057	.057	.059

Table A.8  
 一様分布条件の相関係数の標準偏差 ( $N = 256$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98	1	
m2n2	.062	.062	.062	.062	.063	.067	.074	.087	.108	.147	.181	.197	.218
m2n3	.062	.062	.062	.062	.063	.065	.069	.078	.092	.119	.146	.155	.171
m2n4	.062	.062	.062	.062	.062	.063	.066	.072	.084	.107	.128	.137	.150
m2n5	.062	.062	.062	.061	.061	.062	.064	.068	.080	.099	.118	.127	.138
m2n6	.062	.062	.062	.061	.061	.061	.063	.067	.076	.095	.111	.119	.129
m2n7	.062	.062	.062	.061	.061	.061	.062	.066	.074	.090	.106	.113	.124
m3n3	.062	.062	.062	.062	.063	.064	.067	.071	.080	.098	.115	.124	.137
m3n4	.062	.062	.062	.062	.062	.062	.063	.066	.073	.088	.101	.110	.120
m3n5	.062	.062	.062	.061	.061	.061	.062	.064	.069	.080	.094	.100	.107
m3n6	.062	.062	.062	.061	.061	.060	.060	.062	.067	.077	.087	.093	.102
m3n7	.062	.062	.062	.061	.060	.060	.060	.061	.065	.074	.085	.089	.094
m4n4	.062	.062	.062	.061	.060	.060	.060	.061	.065	.077	.088	.092	.100
m4n5	.062	.062	.061	.061	.060	.059	.058	.058	.060	.069	.080	.084	.090
m4n6	.062	.062	.061	.061	.059	.058	.056	.056	.058	.064	.074	.077	.085
m4n7	.062	.062	.061	.060	.059	.057	.056	.055	.056	.062	.070	.074	.080
m5n5	.062	.062	.061	.060	.059	.057	.056	.055	.056	.062	.071	.074	.080
m5n6	.062	.062	.061	.060	.058	.057	.055	.053	.053	.057	.064	.068	.074
m5n7	.062	.062	.061	.060	.058	.056	.054	.052	.051	.055	.061	.065	.071
m6n6	.062	.062	.061	.060	.058	.056	.053	.051	.050	.053	.058	.061	.066
m6n7	.062	.062	.061	.060	.058	.055	.052	.049	.048	.049	.055	.059	.061
m7n7	.062	.062	.061	.060	.057	.055	.052	.048	.046	.047	.051	.054	.059

Table A.9  
 一様分布条件の相関係数の標準偏差 ( $N = 1024$ )

カテゴリ数	$\rho$											
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98	1
m2n2	.031	.031	.032	.034	.038	.045	.057	.075	.101	.141	.179	.213
m2n3	.031	.031	.032	.034	.038	.044	.052	.065	.086	.114	.141	.168
m2n4	.031	.031	.032	.033	.036	.041	.048	.059	.076	.101	.126	.147
m2n5	.031	.031	.032	.033	.035	.039	.046	.056	.073	.095	.116	.136
m2n6	.031	.031	.032	.033	.035	.038	.044	.054	.068	.091	.107	.127
m2n7	.031	.031	.031	.033	.034	.038	.044	.052	.065	.087	.105	.122
m3n3	.031	.031	.033	.035	.039	.043	.050	.059	.073	.095	.112	.135
m3n4	.031	.031	.032	.034	.037	.041	.047	.054	.066	.084	.101	.115
m3n5	.031	.031	.032	.034	.036	.039	.044	.051	.060	.077	.090	.105
m3n6	.031	.031	.032	.033	.035	.039	.043	.049	.058	.074	.086	.099
m3n7	.031	.031	.032	.033	.035	.038	.042	.048	.056	.070	.082	.095
m4n4	.031	.031	.032	.033	.035	.038	.042	.049	.057	.072	.085	.099
m4n5	.031	.031	.032	.033	.034	.036	.040	.045	.052	.066	.079	.089
m4n6	.031	.031	.031	.032	.034	.035	.038	.042	.049	.061	.071	.083
m4n7	.031	.031	.031	.032	.033	.035	.037	.041	.047	.058	.067	.079
m5n5	.031	.031	.031	.032	.033	.035	.037	.041	.047	.058	.068	.079
m5n6	.031	.031	.031	.032	.032	.033	.036	.039	.044	.054	.064	.073
m5n7	.031	.031	.031	.031	.032	.033	.035	.037	.042	.051	.059	.069
m6n6	.031	.031	.031	.031	.032	.033	.034	.036	.041	.048	.057	.067
m6n7	.031	.031	.031	.031	.031	.032	.033	.035	.039	.046	.053	.062
m7n7	.031	.031	.031	.031	.031	.031	.032	.033	.036	.042	.050	.058

Table A.10  
等距離条件の相関係数の標準偏差 ( $N = 64$ )

カテゴリ数	$\rho$											
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98	1
m2n2	.126	.126	.125	.123	.121	.118	.113	.108	.100	.090	.087	.111
m2n3	.126	.126	.124	.122	.119	.115	.108	.100	.088	.069	.057	.067
m2n4	.126	.126	.124	.122	.118	.114	.106	.097	.084	.064	.046	.039
m2n5	.126	.125	.124	.122	.118	.113	.105	.096	.082	.061	.042	.038
m2n6	.126	.126	.124	.121	.118	.112	.105	.095	.081	.059	.040	.029
m2n7	.126	.125	.124	.121	.117	.112	.104	.094	.080	.058	.039	.028
m3n3	.126	.125	.124	.121	.117	.111	.103	.092	.077	.056	.039	.039
m3n4	.126	.125	.124	.121	.116	.110	.101	.089	.073	.050	.032	.023
m3n5	.126	.125	.123	.120	.116	.109	.099	.088	.071	.048	.029	.018
m3n6	.126	.125	.123	.120	.115	.108	.099	.087	.070	.046	.027	.014
m3n7	.126	.125	.123	.120	.115	.108	.098	.086	.069	.045	.026	.014
m4n4	.126	.125	.123	.120	.115	.108	.099	.086	.070	.046	.027	.019
m4n5	.126	.125	.123	.120	.115	.107	.097	.085	.067	.043	.025	.012
m4n6	.126	.125	.123	.120	.114	.107	.097	.084	.066	.042	.023	.010
m4n7	.126	.125	.123	.119	.114	.107	.096	.083	.065	.041	.022	.008
m5n5	.126	.125	.123	.120	.114	.107	.096	.083	.066	.041	.023	.011
m5n6	.126	.125	.123	.119	.114	.106	.096	.082	.064	.040	.021	.007
m5n7	.126	.125	.123	.119	.114	.106	.095	.082	.064	.039	.020	.006
m6n6	.126	.125	.123	.119	.114	.106	.095	.081	.063	.039	.020	.007
m6n7	.126	.125	.123	.119	.113	.105	.095	.081	.063	.038	.019	.005
m7n7	.126	.125	.123	.119	.113	.105	.094	.080	.062	.037	.018	.005



Table A.11  
等距離条件の相関係数の標準偏差 ( $N = 256$ )

カテゴリ数	$\rho$												
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98	1	
m2n2	.062	.062	.062	.061	.060	.058	.057	.054	.052	.051	.060	.071	.100
m2n3	.062	.062	.061	.061	.059	.057	.054	.050	.044	.036	.037	.044	.061
m2n4	.062	.062	.061	.060	.059	.056	.053	.048	.042	.032	.026	.026	.031
m2n5	.062	.062	.061	.060	.058	.056	.052	.047	.041	.030	.023	.023	.032
m2n6	.062	.062	.061	.060	.058	.055	.052	.047	.040	.029	.021	.019	.020
m2n7	.062	.062	.061	.060	.058	.055	.052	.046	.039	.029	.020	.018	.021
m3n3	.062	.062	.061	.060	.058	.055	.052	.046	.039	.028	.022	.022	.032
m3n4	.062	.062	.061	.060	.057	.054	.050	.044	.036	.025	.017	.015	.020
m3n5	.062	.062	.061	.059	.057	.054	.049	.043	.035	.024	.015	.012	.014
m3n6	.062	.062	.061	.059	.057	.053	.049	.043	.034	.023	.014	.010	.010
m3n7	.062	.062	.061	.059	.057	.053	.049	.042	.034	.022	.013	.010	.011
m4n4	.062	.062	.061	.059	.057	.054	.049	.043	.035	.023	.014	.012	.015
m4n5	.062	.062	.061	.059	.057	.053	.048	.042	.033	.021	.012	.009	.009
m4n6	.062	.062	.061	.059	.056	.053	.048	.041	.033	.021	.011	.008	.007
m4n7	.062	.062	.061	.059	.056	.053	.048	.041	.032	.020	.011	.007	.006
m5n5	.062	.062	.061	.059	.056	.053	.048	.041	.033	.020	.011	.008	.008
m5n6	.062	.062	.061	.059	.056	.052	.047	.041	.032	.020	.010	.007	.005
m5n7	.062	.062	.061	.059	.056	.052	.047	.040	.031	.019	.010	.006	.004
m6n6	.062	.062	.061	.059	.056	.052	.047	.040	.031	.019	.010	.006	.005
m6n7	.062	.062	.061	.059	.056	.052	.047	.040	.031	.019	.009	.006	.003
m7n7	.062	.062	.061	.059	.056	.052	.047	.040	.030	.018	.009	.005	.003

Table A.12  
等距離条件の相関係数の標準偏差 ( $N = 1024$ )

カテゴリ数	$\rho$											
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.98	1
m2n2	.031	.031	.031	.030	.030	.029	.029	.028	.029	.035	.052	.096
m2n3	.031	.031	.031	.030	.030	.029	.028	.026	.023	.020	.030	.059
m2n4	.031	.031	.030	.030	.029	.028	.027	.024	.021	.017	.018	.029
m2n5	.031	.031	.030	.030	.029	.028	.026	.024	.020	.015	.015	.030
m2n6	.031	.031	.030	.030	.029	.028	.026	.023	.020	.015	.013	.018
m2n7	.031	.031	.030	.030	.029	.027	.026	.023	.019	.014	.011	.018
m3n3	.031	.031	.031	.030	.029	.028	.027	.024	.021	.016	.015	.030
m3n4	.031	.031	.030	.030	.029	.027	.025	.023	.019	.013	.010	.019
m3n5	.031	.031	.030	.029	.028	.027	.025	.022	.018	.012	.008	.013
m3n6	.031	.031	.030	.029	.028	.027	.024	.021	.017	.011	.007	.009
m3n7	.031	.031	.030	.029	.028	.026	.024	.021	.017	.011	.007	.009
m4n4	.031	.031	.030	.030	.029	.027	.025	.022	.018	.012	.008	.014
m4n5	.031	.031	.030	.029	.028	.026	.024	.021	.017	.011	.007	.008
m4n6	.031	.031	.030	.029	.028	.026	.024	.021	.016	.010	.006	.007
m4n7	.031	.031	.030	.029	.028	.026	.024	.020	.016	.010	.005	.005
m5n5	.031	.031	.030	.029	.028	.026	.024	.021	.016	.010	.006	.008
m5n6	.031	.031	.030	.029	.028	.026	.024	.020	.016	.010	.005	.004
m5n7	.031	.031	.030	.029	.028	.026	.023	.020	.016	.009	.005	.004
m6n6	.031	.031	.030	.029	.028	.026	.023	.020	.016	.009	.005	.005
m6n7	.031	.031	.030	.029	.028	.026	.023	.020	.015	.009	.005	.003
m7n7	.031	.031	.030	.029	.028	.026	.023	.020	.015	.009	.004	.003

付録 B

研究 2-a の補足資料

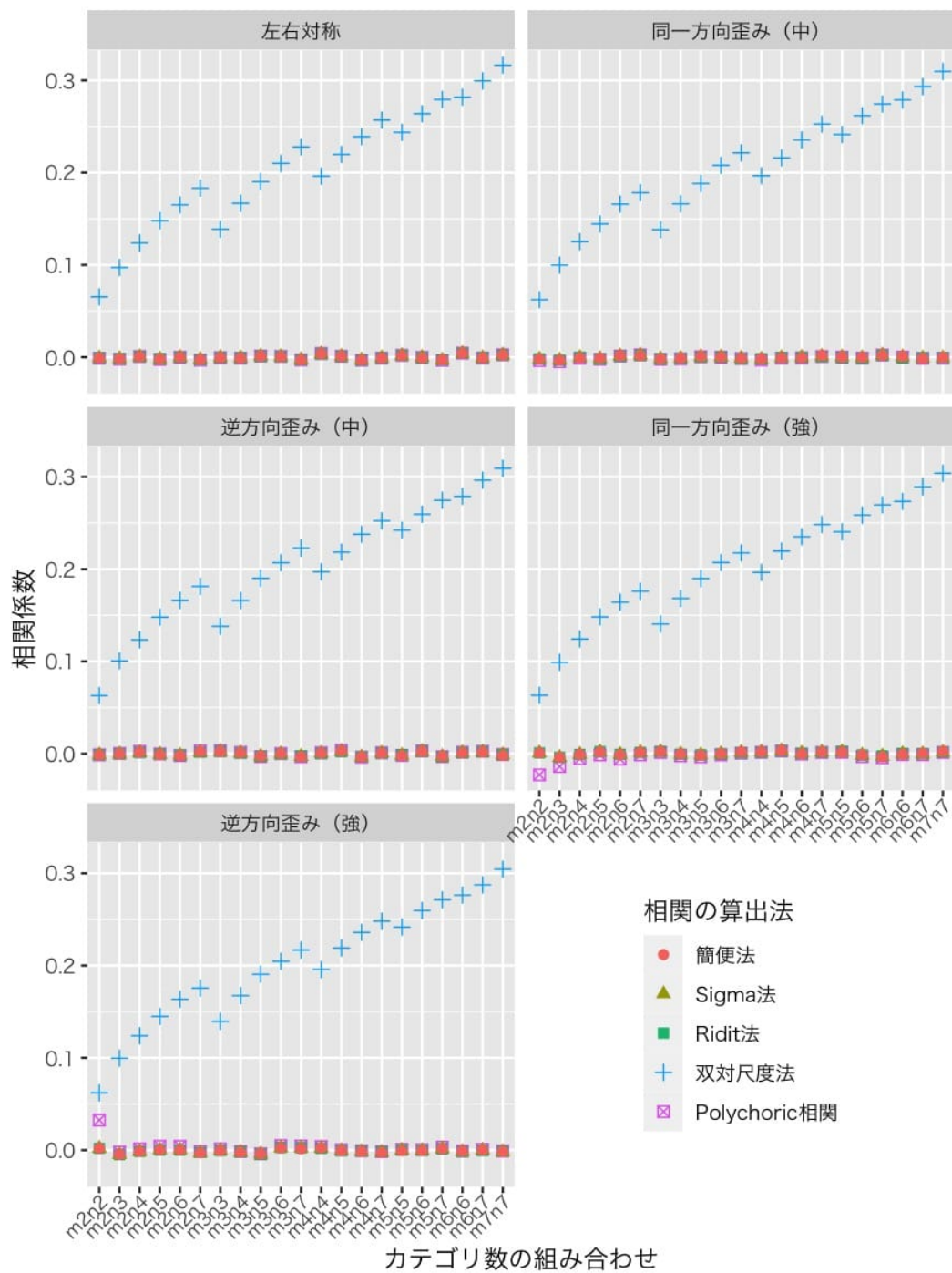


Figure B.1.  $\rho = 0$  の場合の平均相関係数

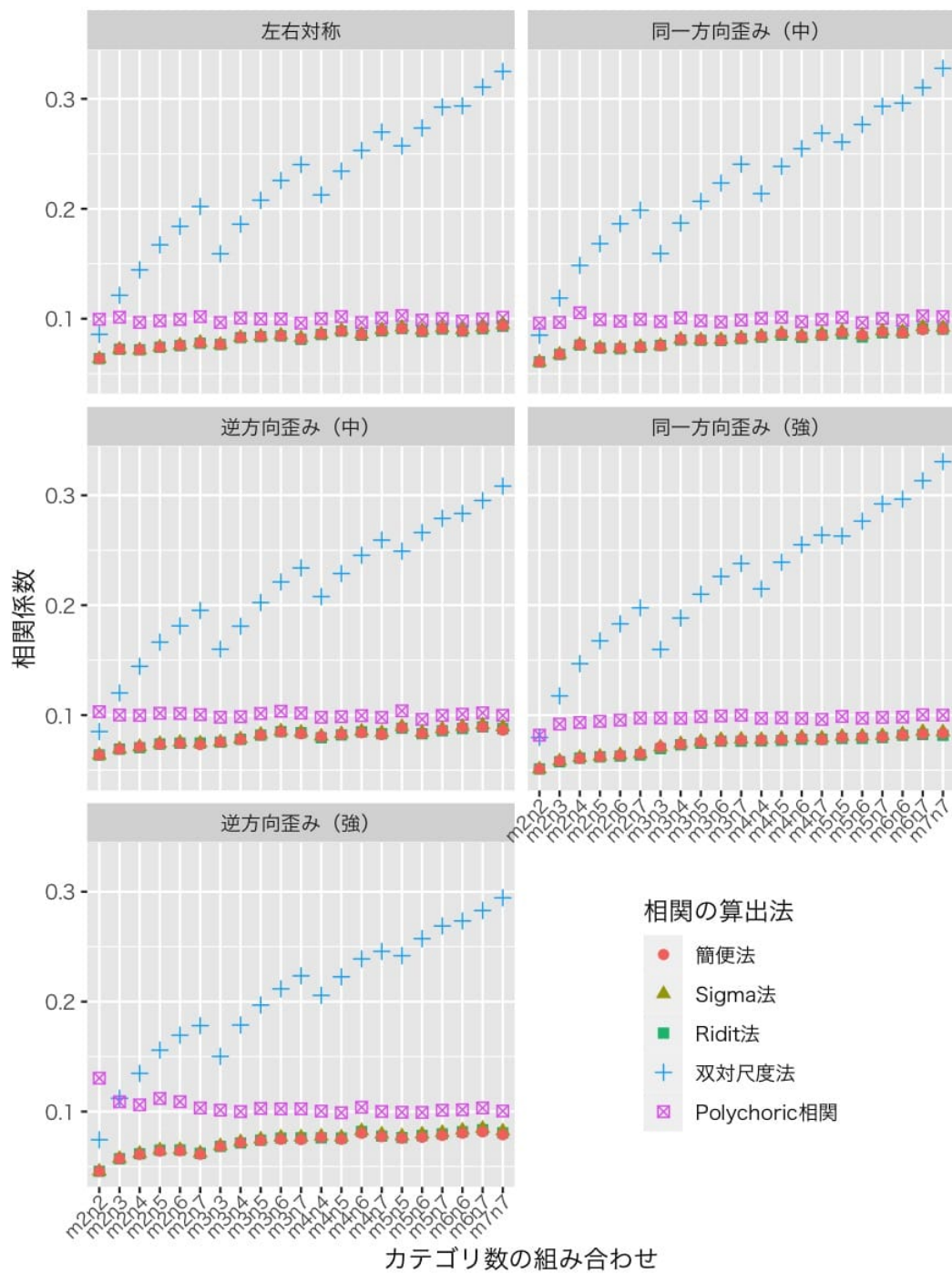


Figure B.2.  $\rho = 0.1$  の場合の平均相関係数

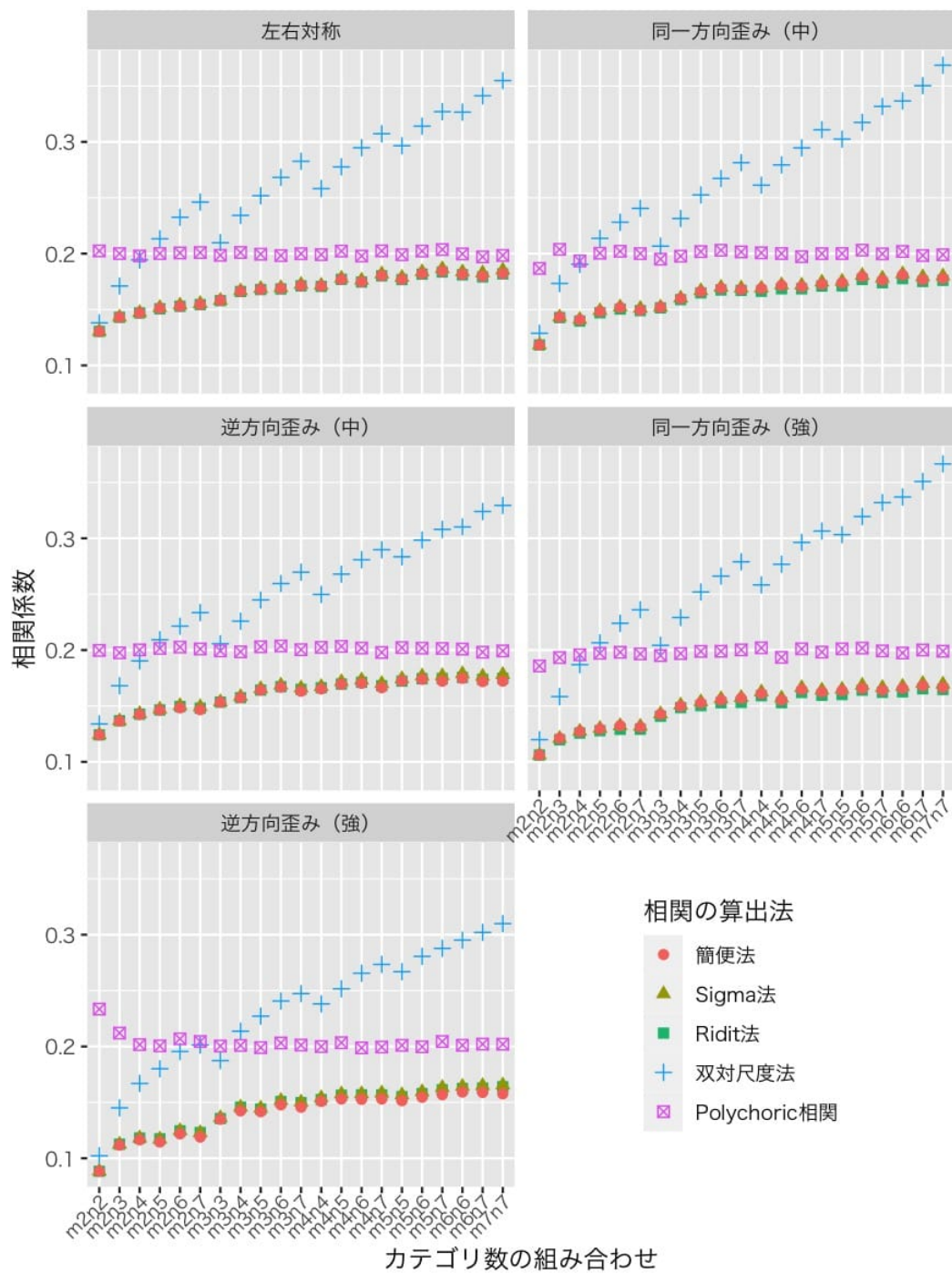


Figure B.3.  $\rho = 0.2$  の場合の平均相関係数

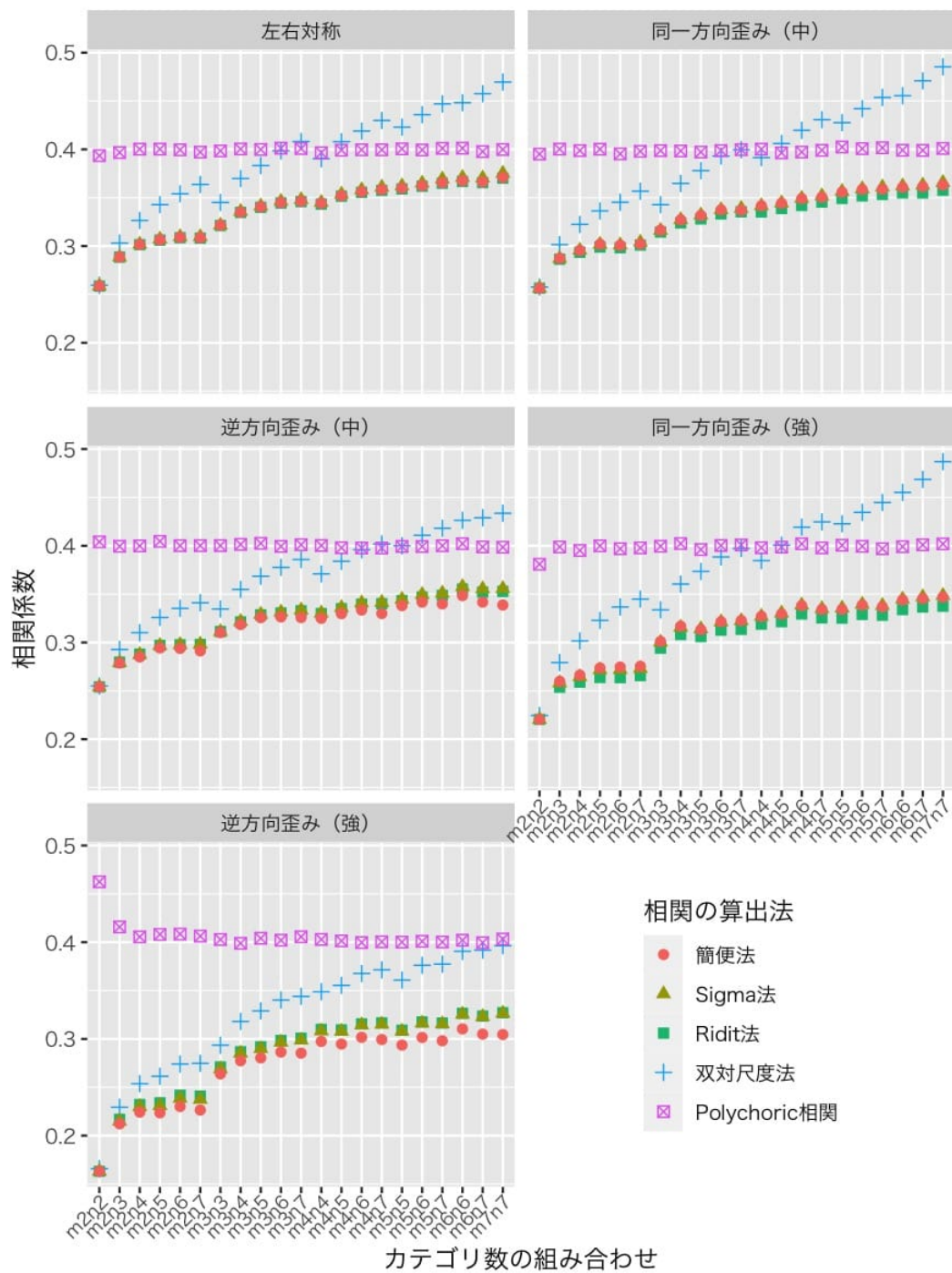


Figure B.4.  $\rho = 0.4$  の場合の平均相関係数

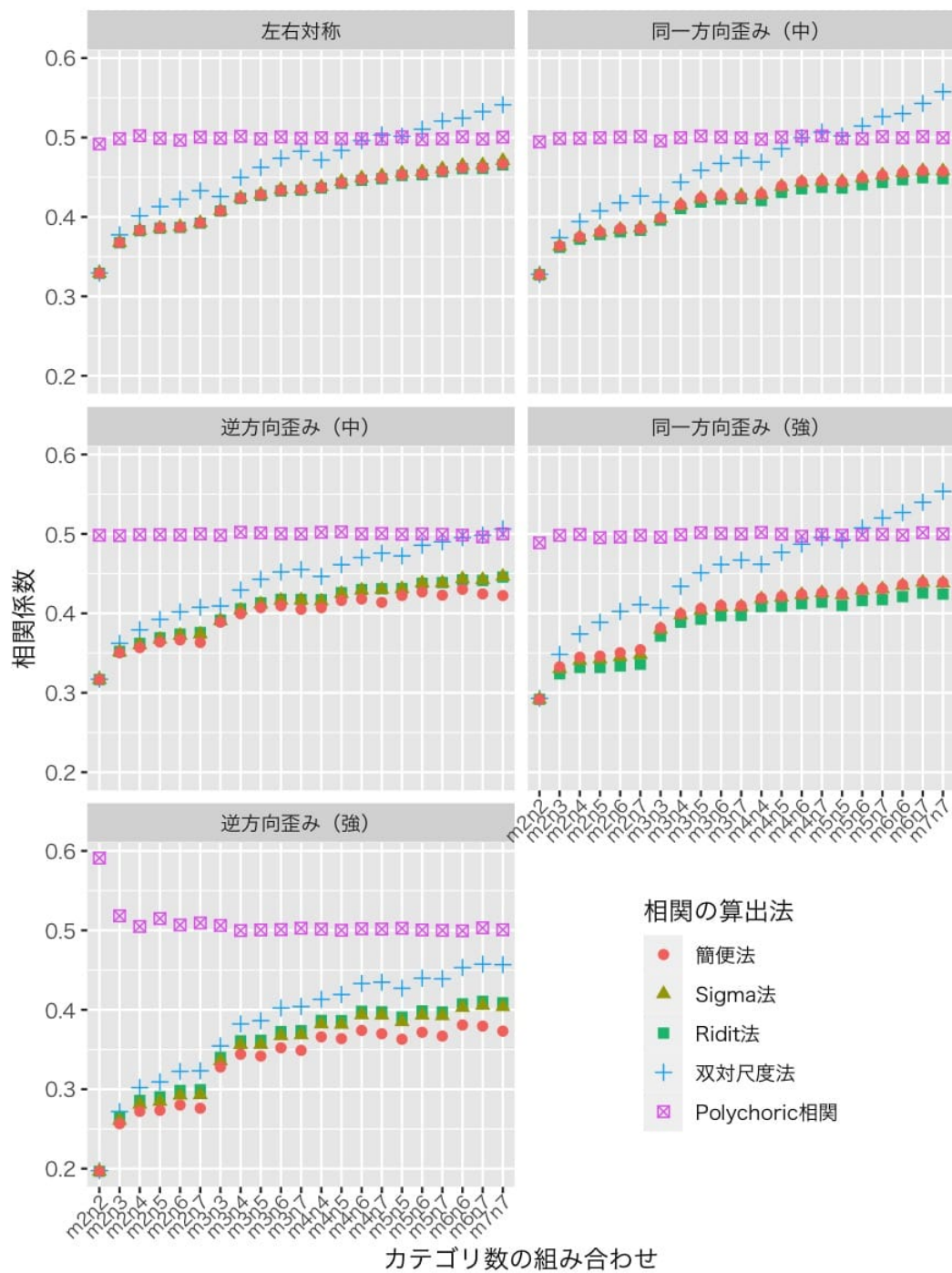


Figure B.5.  $\rho = 0.5$  の場合の平均相関係数



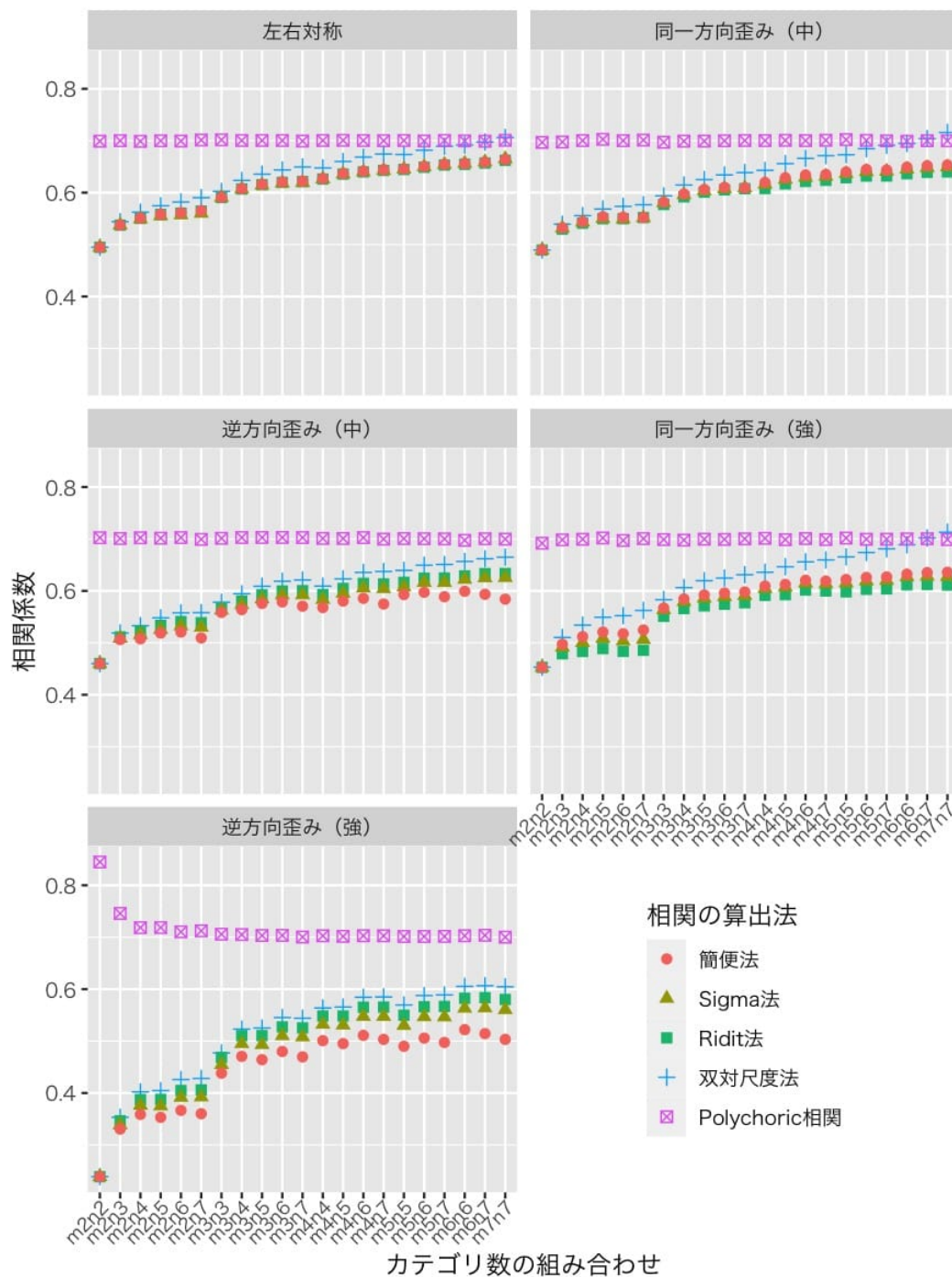


Figure B.6.  $\rho = 0.7$  の場合の平均相関係数

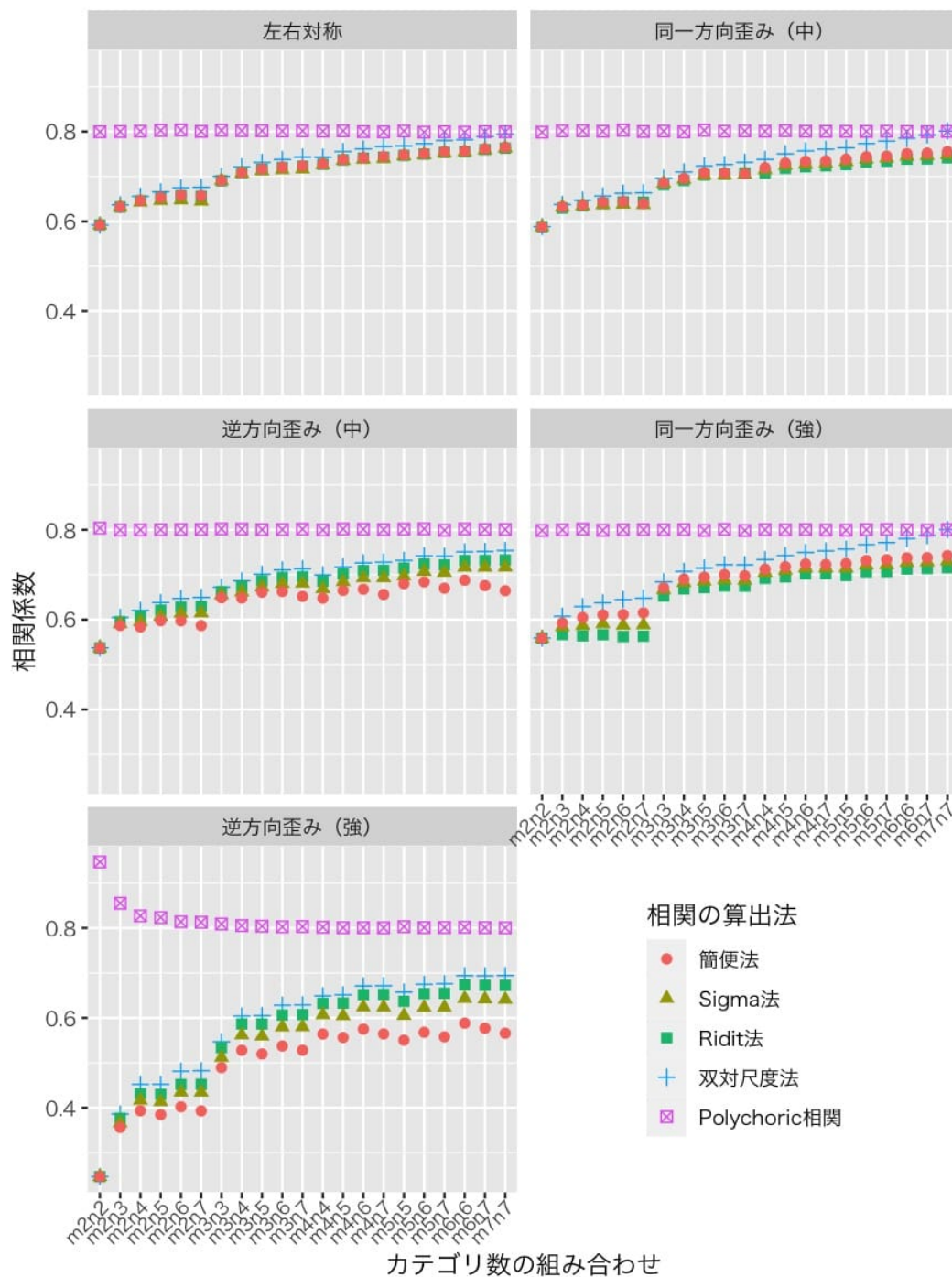


Figure B.7.  $\rho = 0.8$  の場合の平均相関係数

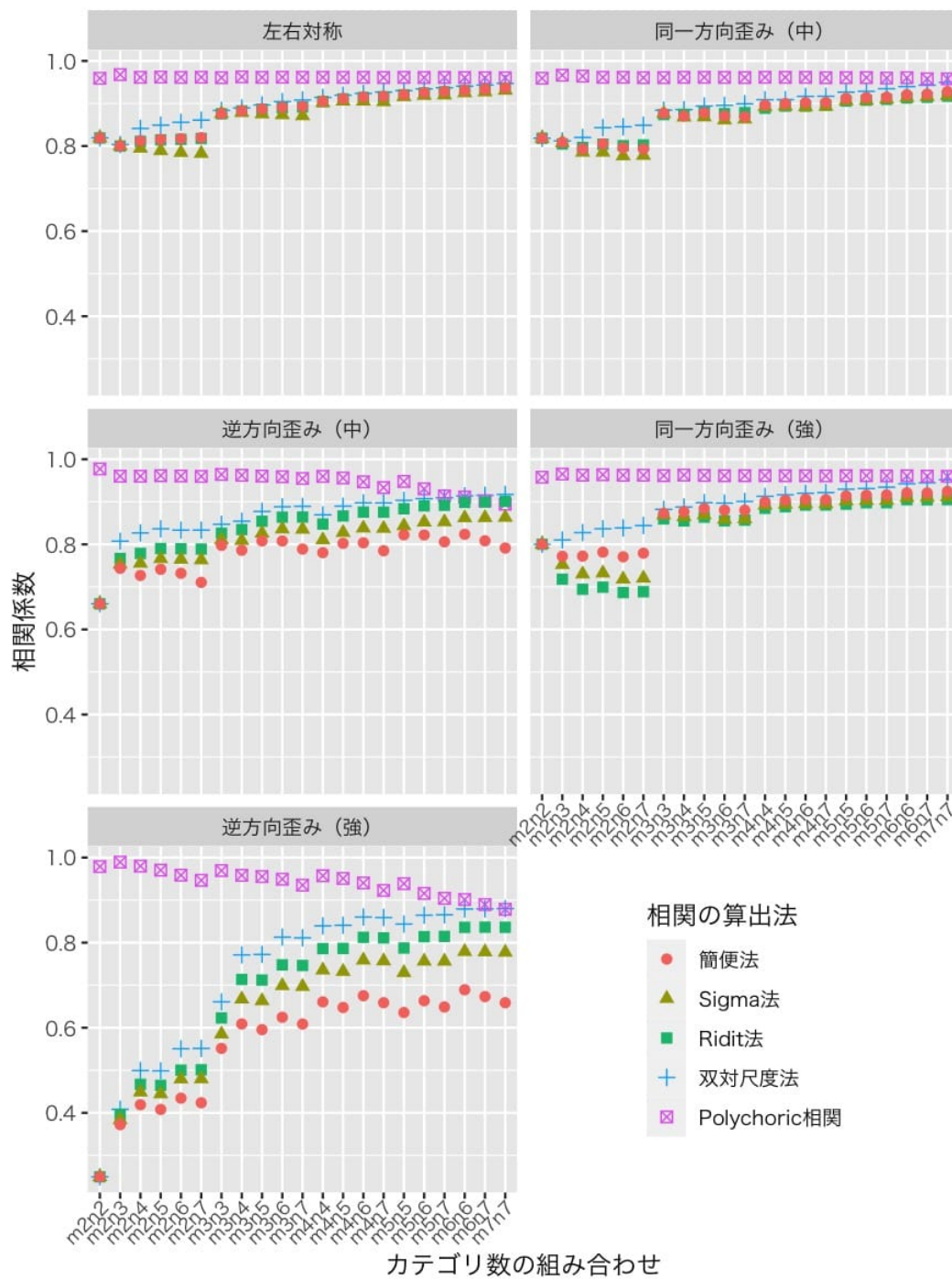


Figure B.8.  $\rho = 0.96$  の場合の平均相関係数

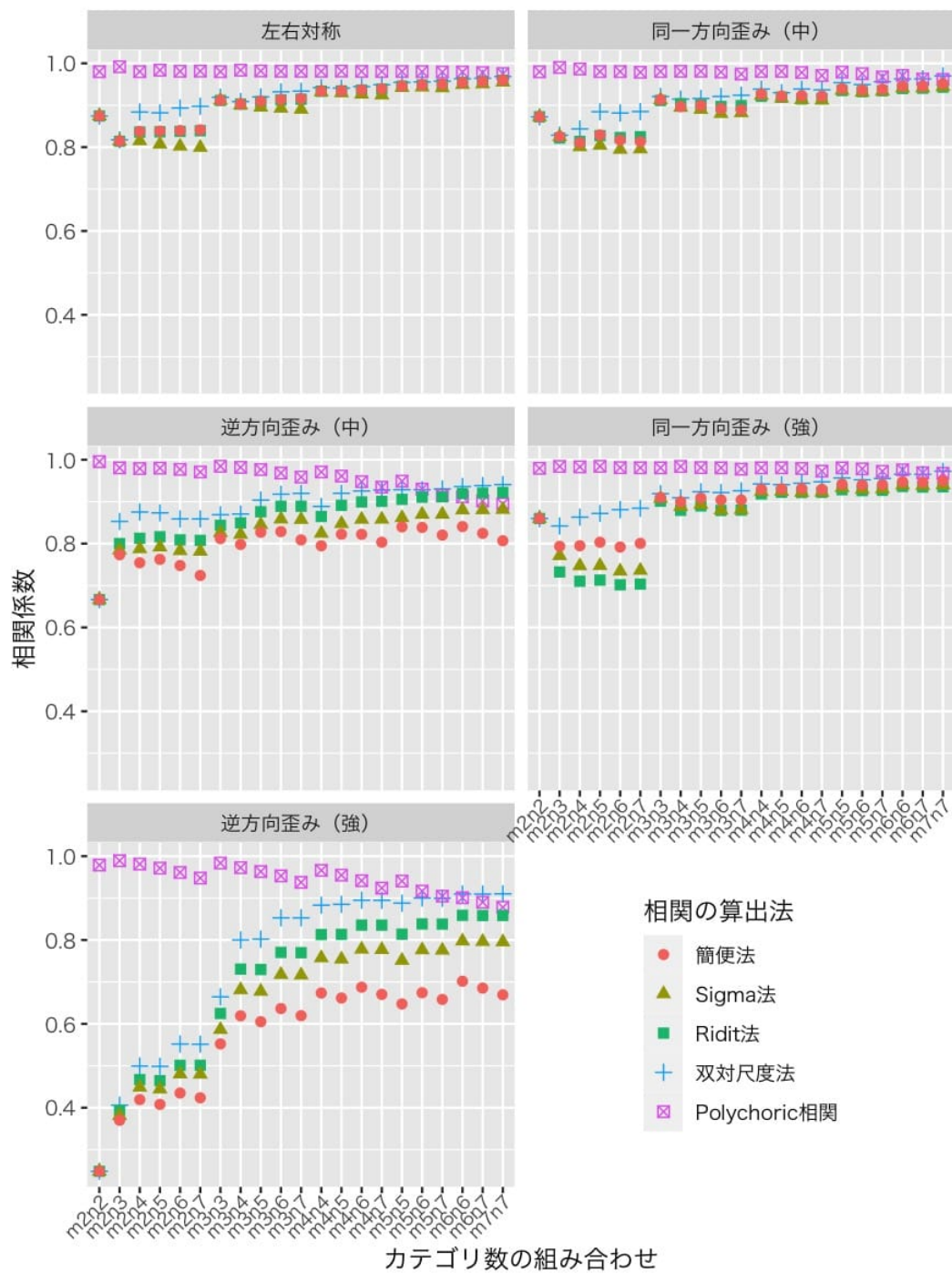


Figure B.9.  $\rho = 0.98$  の場合の平均相関係数

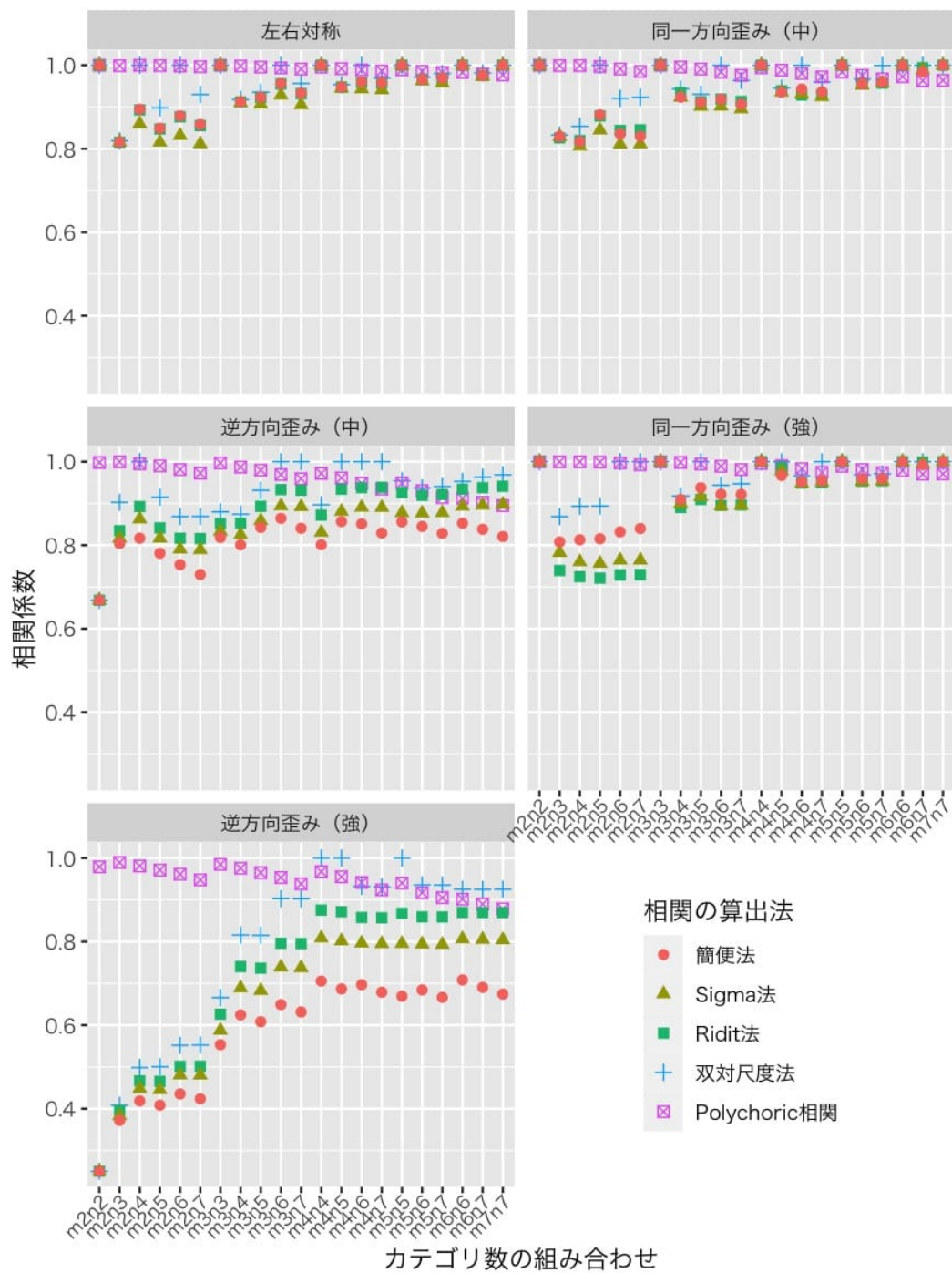


Figure B.10.  $\rho = 1.0$  の場合の平均相関係数

付録 C

研究 3 の補足資料

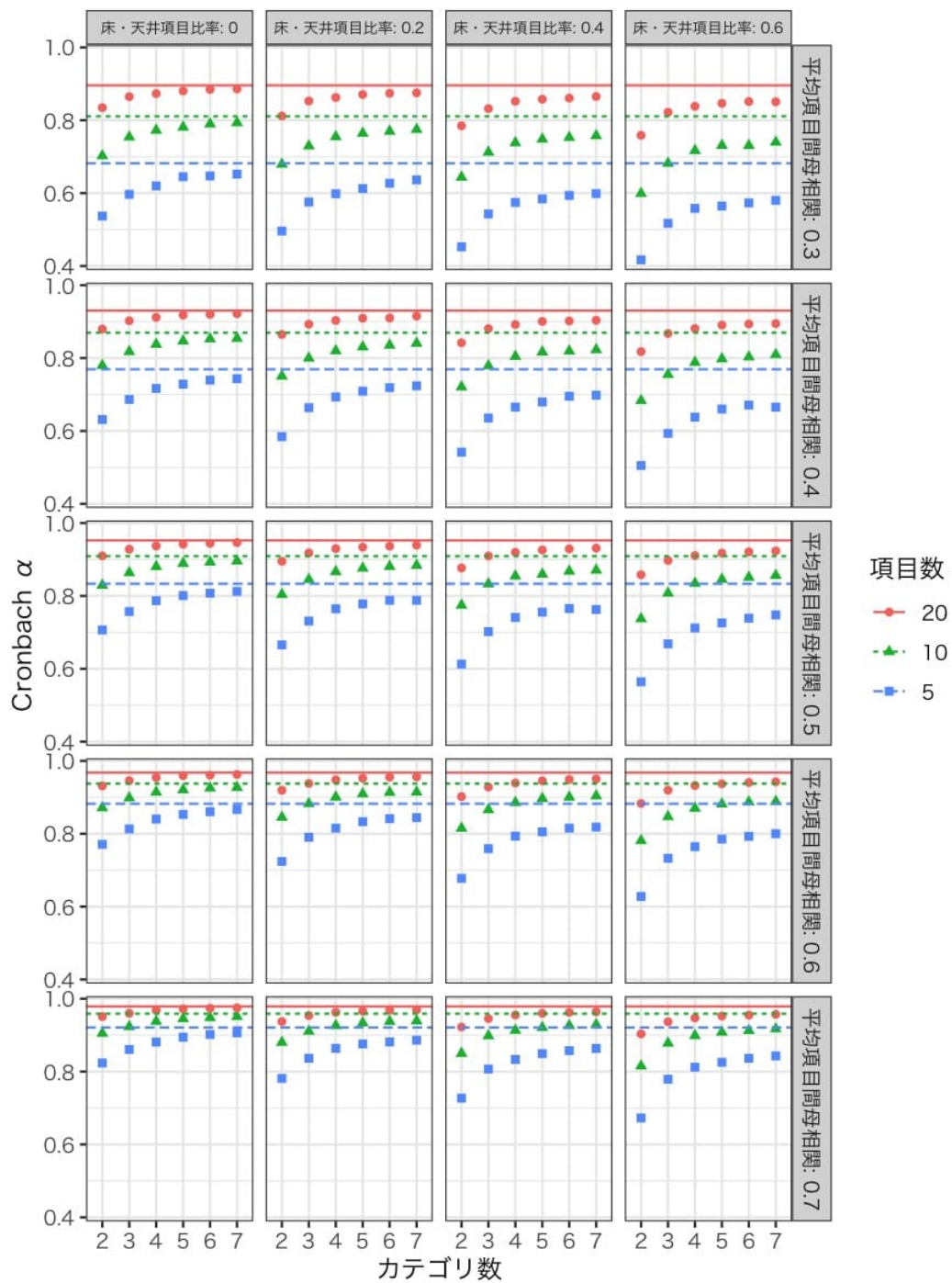


Figure C.1. サンプルサイズが 50 の場合の  $\alpha$  の平均

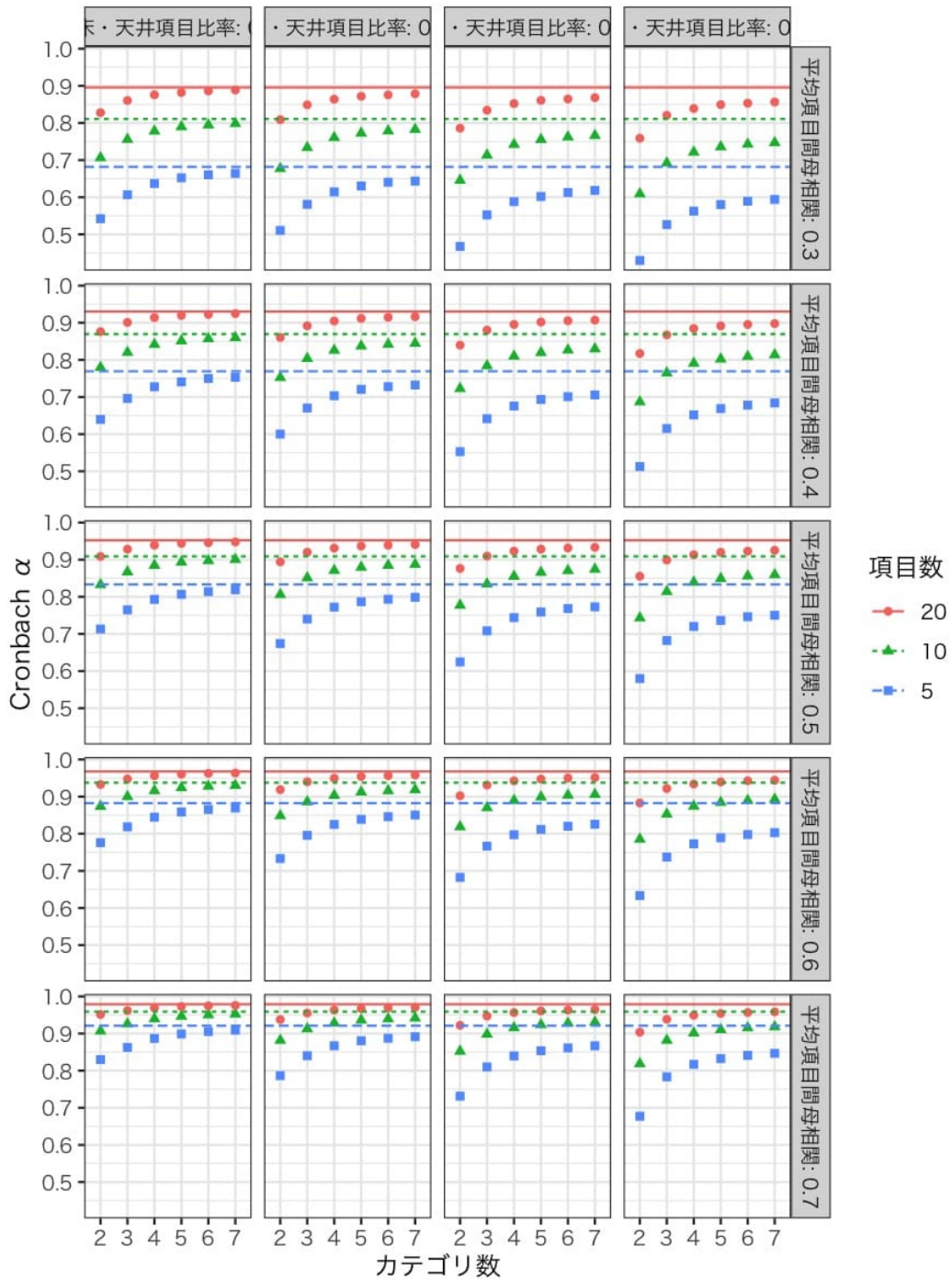


Figure C.2. サンプルサイズが 100 の場合の  $\alpha$  の平均



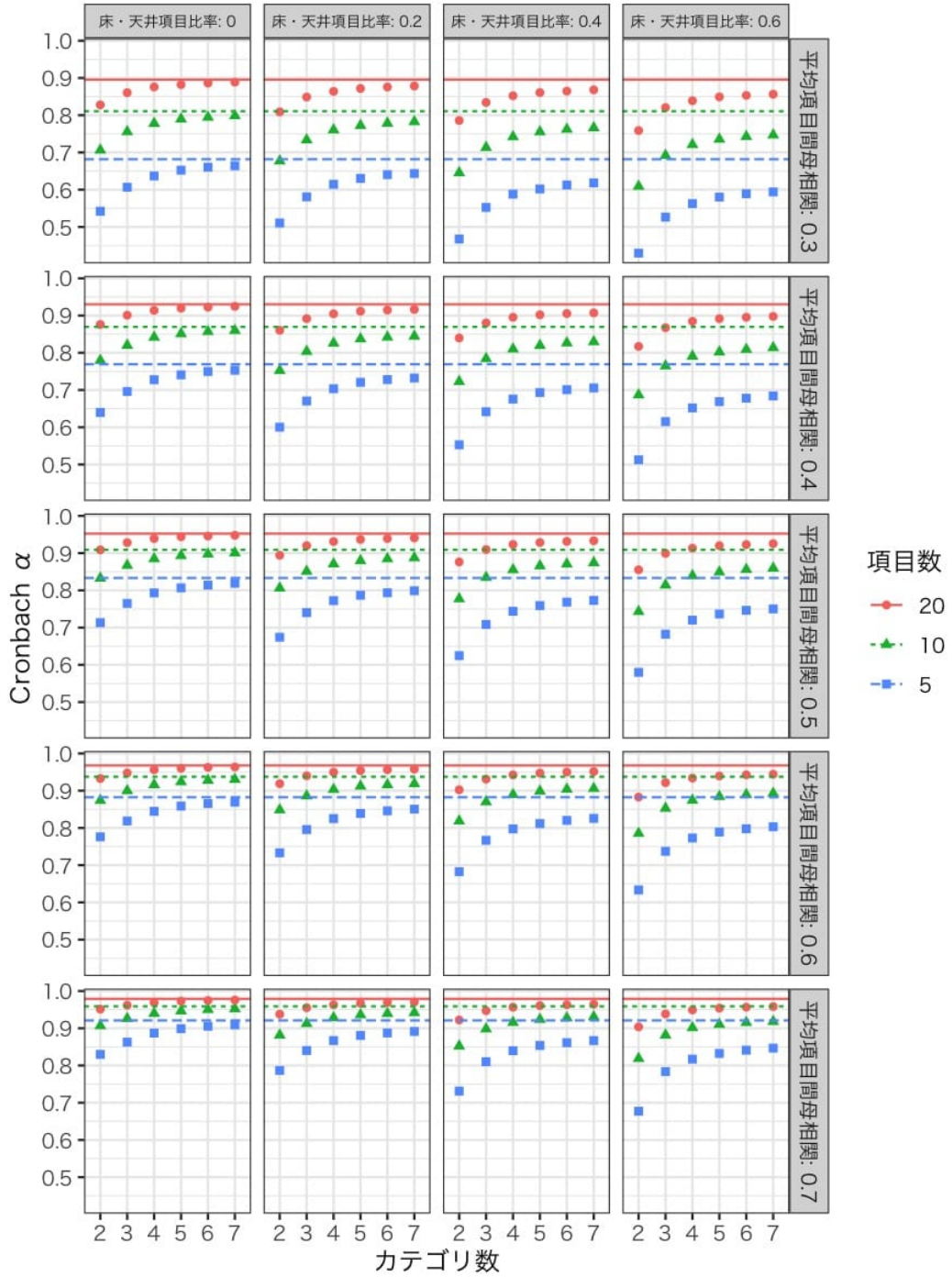


Figure C.3. サンプルサイズが 500 の場合の  $\alpha$  の平均

# 付録 D

## 研究 4-a の補足資料

Table D.1

近似母集団研究における信頼性係数とバイアス

関値の均質性	モデル A					モデル B					モデル C				
	$\rho_{correct}$	$\rho_{mis1}$	$\rho_{mis2}$	$\alpha$	$\omega$	$\rho_{correct}$	$\rho_{mis1}$	$\rho_{mis2}$	$\alpha$	$\omega$	$\rho_{correct}$	$\rho_{mis1}$	$\rho_{mis2}$	$\alpha$	$\omega$
2 カテゴリ															
均質	.845	.692	.571	.802	.845	.878	.881	.886	.837	.876	.838	.870	.906	.867	.899
	(.000)	(-.153)	(-.274)	(-.043)	(.000)	(.000)	(.003)	(.007)	(-.041)	(-.002)	(.000)	(.033)	(.069)	(.030)	(.061)
中程度に不均質	.840	.687	.566	.794	.835	.874	.877	.881	.830	.867	.829	.865	.897	.858	.889
	(.000)	(-.152)	(-.274)	(-.046)	(-.004)	(.000)	(.003)	(.007)	(-.044)	(-.007)	(.000)	(.037)	(.069)	(.030)	(.061)
かなり不均質	.831	.679	.559	.783	.823	.866	.868	.873	.818	.854	.834	.863	.896	.853	.883
	(-.000)	(-.152)	(-.273)	(-.049)	(-.009)	(.000)	(.003)	(.008)	(-.048)	(-.011)	(.000)	(.030)	(.062)	(.020)	(.049)
極めて不均質	.731	.606	.491	.653	.690	.799	.802	.799	.716	.745	.709	.746	.825	.771	.797
	(-.001)	(-.126)	(-.241)	(-.079)	(-.042)	(.001)	(.003)	(.000)	(-.083)	(-.053)	(.001)	(.038)	(.116)	(.062)	(.089)
3 カテゴリ															
均質	.883	.720	.590	.840	.883	.908	.912	.919	.870	.908	.865	.900	.936	.895	.926
	(.000)	(-.162)	(-.293)	(-.043)	(.000)	(-.000)	(.003)	(.010)	(-.038)	(-.001)	(.000)	(.035)	(.071)	(.030)	(.061)
中程度に不均質	.878	.717	.587	.834	.878	.906	.909	.914	.866	.904	.863	.895	.935	.892	.923
	(-.000)	(-.162)	(-.291)	(-.044)	(-.000)	(.000)	(.004)	(.009)	(-.040)	(-.001)	(.000)	(.032)	(.072)	(.029)	(.060)
かなり不均質	.877	.717	.587	.831	.874	.906	.909	.915	.863	.900	.862	.897	.934	.890	.920
	(-.000)	(-.161)	(-.290)	(-.047)	(-.003)	(.000)	(.003)	(.010)	(-.043)	(-.005)	(.000)	(.035)	(.072)	(.028)	(.058)
4 カテゴリ															
均質	.900	.735	.603	.856	.900	.924	.928	.936	.885	.924	.876	.911	.949	.906	.938
	(.000)	(-.165)	(-.297)	(-.044)	(-.000)	(.000)	(.004)	(.012)	(-.039)	(-.000)	(.000)	(.036)	(.073)	(.031)	(.062)
中程度に不均質	.894	.730	.599	.848	.892	.919	.922	.929	.877	.916	.868	.906	.942	.899	.930
	(-.000)	(-.164)	(-.296)	(-.047)	(-.002)	(-.000)	(.003)	(.010)	(-.042)	(-.003)	(-.000)	(.037)	(.074)	(.031)	(.062)
かなり不均質	.887	.724	.594	.839	.881	.914	.917	.925	.868	.906	.871	.904	.941	.894	.924
	(.000)	(-.163)	(-.293)	(-.049)	(-.006)	(.000)	(.003)	(.011)	(-.045)	(-.008)	(.000)	(.034)	(.070)	(.024)	(.054)
5 カテゴリ															
均質	.906	.739	.605	.862	.906	.929	.932	.941	.890	.928	.881	.917	.954	.911	.942
	(.000)	(-.167)	(-.301)	(-.044)	(.000)	(.000)	(.004)	(.013)	(-.038)	(-.000)	(-.000)	(.036)	(.073)	(.030)	(.062)
中程度に不均質	.895	.730	.598	.849	.895	.920	.924	.931	.878	.918	.870	.906	.944	.900	.931
	(-.000)	(-.165)	(-.297)	(-.047)	(-.001)	(-.000)	(.004)	(.010)	(-.042)	(-.002)	(-.000)	(.035)	(.074)	(.030)	(.061)
かなり不均質	.887	.723	.592	.836	.880	.915	.918	.925	.867	.904	.871	.906	.943	.894	.923
	(.000)	(-.164)	(-.295)	(-.051)	(-.007)	(-.000)	(.003)	(.010)	(-.049)	(-.011)	(.000)	(.035)	(.072)	(.023)	(.052)
6 カテゴリ															
均質	.910	.741	.606	.866	.910	.931	.935	.944	.893	.931	.883	.919	.957	.914	.945
	(.000)	(-.168)	(-.303)	(-.044)	(-.000)	(-.000)	(.004)	(.013)	(-.038)	(-.000)	(.000)	(.036)	(.074)	(.031)	(.062)
中程度に不均質	.899	.733	.600	.854	.903	.923	.927	.933	.882	.925	.877	.908	.952	.906	.938
	(.000)	(-.166)	(-.299)	(-.045)	(.004)	(-.000)	(.004)	(.010)	(-.041)	(.001)	(-.000)	(.030)	(.074)	(.029)	(.060)
かなり不均質	.896	.730	.598	.847	.893	.923	.926	.934	.878	.917	.874	.911	.951	.903	.933
	(-.000)	(-.166)	(-.298)	(-.049)	(-.003)	(.000)	(.003)	(.011)	(-.045)	(-.006)	(.001)	(.038)	(.077)	(.030)	(.060)

Table D.2  
不適解の発生率 (2 カテゴリ)

サンプルサイズ	モデル A				モデル B				モデル C			
	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$
均質												
50	34.0%	30.2%	25.6%	0.0%	60.2%	64.2%	57.6%	0.2%	100.0%	100.0%	100.0%	0.4%
100	5.0%	4.2%	3.8%	0.0%	27.4%	41.6%	28.6%	0.0%	18.2%	12.6%	31.2%	0.0%
200	0.0%	0.0%	0.0%	0.0%	8.0%	25.2%	8.4%	0.0%	2.0%	0.6%	5.4%	0.2%
500	0.0%	0.0%	0.0%	0.0%	0.2%	14.4%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	6.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
中程度に不均質												
50	31.2%	34.8%	29.8%	0.2%	67.2%	70.6%	66.0%	0.0%	100.0%	100.0%	100.0%	0.8%
100	4.6%	2.6%	4.2%	0.2%	39.0%	47.8%	36.2%	0.0%	37.0%	27.8%	41.6%	0.4%
200	0.2%	0.2%	0.0%	0.0%	15.0%	30.4%	15.0%	0.0%	5.6%	1.6%	14.2%	0.2%
500	0.0%	0.0%	0.0%	0.0%	1.2%	16.2%	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	6.6%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
かなり不均質												
50	39.0%	40.2%	34.2%	0.2%	65.8%	64.2%	60.2%	0.2%	100.0%	100.0%	100.0%	0.4%
100	11.0%	9.0%	8.0%	0.0%	47.0%	53.4%	39.8%	0.0%	45.0%	38.6%	44.0%	1.0%
200	0.8%	1.0%	1.4%	0.0%	24.0%	40.6%	19.4%	0.0%	6.0%	3.4%	11.0%	0.4%
500	0.0%	0.0%	0.0%	0.0%	3.8%	22.4%	4.0%	0.0%	0.0%	0.0%	0.6%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.2%	14.4%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
極めて不均質												
50	95.4%	95.8%	99.0%	1.4%	99.4%	99.4%	97.8%	1.4%	100.0%	100.0%	100.0%	1.6%
100	40.4%	48.6%	62.0%	1.0%	63.8%	54.6%	49.8%	0.8%	100.0%	96.8%	98.6%	4.4%
200	4.6%	7.6%	13.2%	0.4%	18.4%	22.8%	12.8%	0.2%	29.8%	49.8%	68.8%	5.0%
500	2.4%	2.4%	3.2%	0.0%	11.0%	25.0%	10.8%	0.0%	6.4%	12.4%	7.8%	0.2%
1000	1.4%	1.8%	1.8%	0.0%	5.0%	27.6%	9.4%	0.0%	2.2%	4.6%	1.0%	0.0%

Table D.3  
不適解の発生率 (3 カテゴリ)

サンプルサイズ	モデル A			モデル B			モデル C					
	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$
均質												
50	13.2%	11.4%	9.8%	0.0%	100.0%	100.0%	45.8%	0.2%	100.0%	100.0%	100.0%	0.6%
100	0.2%	0.0%	0.0%	0.0%	15.6%	34.0%	16.8%	0.0%	100.0%	100.0%	10.8%	0.0%
200	0.0%	0.0%	0.0%	0.0%	2.6%	20.2%	3.6%	0.0%	0.2%	0.0%	1.0%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	8.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	2.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
中程度に不均質												
50	11.2%	11.6%	8.6%	0.0%	100.0%	99.6%	51.0%	0.0%	100.0%	100.0%	100.0%	0.4%
100	0.8%	0.8%	0.6%	0.0%	16.4%	33.4%	14.6%	0.0%	100.0%	100.0%	14.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	2.2%	19.0%	3.4%	0.0%	0.0%	0.0%	1.2%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	7.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
かなり不均質												
50	14.4%	12.2%	11.6%	0.0%	100.0%	100.0%	50.6%	0.0%	100.0%	100.0%	100.0%	0.2%
100	0.6%	0.2%	0.6%	0.0%	18.0%	35.2%	19.2%	0.0%	100.0%	100.0%	12.2%	0.2%
200	0.0%	0.0%	0.0%	0.0%	3.2%	21.6%	3.4%	0.0%	0.2%	0.2%	1.2%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	7.4%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table D.4  
不適解の発生率 (4 カテゴリー)

サンプルサイズ	モデル A			モデル B			モデル C					
	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$
均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.6%
100	0.2%	0.2%	0.0%	0.0%	4.2%	19.2%	5.4%	0.0%	100.0%	100.0%	100.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	0.2%	7.4%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
中程度に不均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%
100	0.0%	0.0%	0.0%	0.0%	9.4%	23.6%	7.2%	0.0%	100.0%	100.0%	100.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	0.2%	13.8%	0.4%	0.0%	0.0%	0.0%	0.4%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
かなり不均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.6%
100	0.2%	0.6%	0.4%	0.0%	14.6%	28.2%	14.4%	0.0%	100.0%	100.0%	100.0%	0.2%
200	0.0%	0.0%	0.0%	0.0%	2.6%	17.6%	1.8%	0.0%	0.2%	0.0%	0.8%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	4.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table D.5  
不適解の発生率 (5 カテゴリ)

サンプルサイズ	モデル A			モデル B			モデル C					
	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$
均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.2%
100	0.0%	0.0%	0.0%	0.0%	2.6%	18.0%	3.2%	0.0%	100.0%	100.0%	100.0%	0.2%
200	0.0%	0.0%	0.0%	0.0%	0.2%	9.6%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
中程度に不均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.2%	100.0%	100.0%	100.0%	0.6%
100	0.0%	0.0%	0.0%	0.0%	6.0%	20.0%	7.4%	0.0%	100.0%	100.0%	100.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	0.2%	9.8%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	2.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
かなり不均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	1.0%
100	0.2%	0.0%	0.0%	0.0%	13.6%	30.6%	10.4%	0.0%	100.0%	100.0%	100.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	2.8%	16.0%	2.2%	0.0%	0.0%	0.0%	1.0%	0.2%
500	0.0%	0.0%	0.0%	0.0%	0.0%	4.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table D.6  
不適解の発生率 (6 カテゴリー)

サンプルサイズ	モデル A			モデル B			モデル C					
	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$	正しく特定	誤特定 1	誤特定 2	$\omega$
	均質											
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.4%
100	10.8%	10.8%	10.8%	0.0%	22.2%	35.0%	22.4%	0.0%	100.0%	100.0%	100.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	0.0%	6.4%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
中程度に不均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.4%
100	11.8%	11.8%	11.8%	0.0%	18.6%	31.4%	18.2%	0.0%	100.0%	100.0%	100.0%	0.0%
200	0.0%	0.0%	0.0%	0.0%	0.2%	10.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	1.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
かなり不均質												
50	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.0%	100.0%	100.0%	100.0%	0.4%
100	7.2%	7.0%	7.0%	0.0%	17.4%	29.4%	17.8%	0.0%	100.0%	100.0%	100.0%	0.2%
200	0.0%	0.0%	0.0%	0.0%	0.4%	11.2%	0.6%	0.0%	0.0%	0.0%	0.8%	0.0%
500	0.0%	0.0%	0.0%	0.0%	0.0%	2.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1000	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table D.7  
 サンプルにおける平均信頼性係数とバイアス (モデル A)

	均質										中程度に不均質										かなり不均質										極めて不均質																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
	50		100		200		500		1000		50		100		200		500		1000		50		100		200		500		1000		50		100		200		500		1000																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$	$T_{correct}$	$T_{bias1}$	$T_{bias2}$	$\alpha$	$\omega$																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
2 カテゴリ																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
$T_{correct}$	.898	.872	.860	.850	.848	.897	.871	.855	.845	.843	.886	.865	.849	.838	.835	.446	.644	.704	.732	$T_{correct}$	.918	.892	.886	.882	.880	.882	.882	.882	.880	.882	.882	.882	.882	.880	.882	.882	.882	.882	$T_{correct}$	.918	.892	.886	.882	.880	.882	.882	.882	.880	.882	.882	.882	.882	.880	.882	.882	.882	.882	.882	$T_{correct}$	.918	.892	.886	.882	.880	.882	.882	.882	.880	.882	.882	.882	.882	.880	.882	.882	.882	.882	.882	$T_{correct}$	.918	.892	.886	.882	.880	.882	.882	.882	.880	.882	.882	.882	.882	.880	.882	.882	.882	.882	.882	$T_{correct}$	.918	.892	.886	.882	.880	.882	.882	.882	.880	.882	.882	.882	.882	.880	.882	.882	.882	.882	.882	$T_{bias1}$	.754	.715	.704	.696	.695	.737	.709	.700	.692	.690	.722	.705	.697	.686	.682	.440	.536	.582	.606	$T_{bias1}$	.715	.676	.665	.657	.656	.709	.681	.672	.664	.662	.696	.679	.670	.660	.652	.440	.536	.582	.606	$T_{bias1}$	.715	.676	.665	.657	.656	.709	.681	.672	.664	.662	.696	.679	.670	.660	.652	.440	.536	.582	.606	$T_{bias1}$	.715	.676	.665	.657	.656	.709	.681	.672	.664	.662	.696	.679	.670	.660	.652	.440	.536	.582	.606	$T_{bias1}$	.715	.676	.665	.657	.656	.709	.681	.672	.664	.662	.696	.679	.670	.660	.652	.440	.536	.582	.606	$T_{bias2}$	.801	.768	.757	.749	.748	.801	.771	.762	.753	.751	.783	.770	.761	.752	.743	.444	.647	.693	.717	$T_{bias2}$	.768	.729	.718	.710	.709	.768	.738	.729	.720	.718	.750	.737	.728	.719	.710	.444	.647	.693	.717	$T_{bias2}$	.768	.729	.718	.710	.709	.768	.738	.729	.720	.718	.750	.737	.728	.719	.710	.444	.647	.693	.717	$T_{bias2}$	.768	.729	.718	.710	.709	.768	.738	.729	.720	.718	.750	.737	.728	.719	.710	.444	.647	.693	.717	$T_{bias2}$	.768	.729	.718	.710	.709	.768	.738	.729	.720	.718	.750	.737	.728	.719	.710	.444	.647	.693	.717	$\alpha$	(-.051)	(-.048)	(-.044)	(-.043)	(-.043)	(-.042)	(-.041)	(-.040)	(-.039)	(-.039)	(-.038)	(-.038)	(-.038)	(-.038)	(-.038)	(-.037)	(-.037)	(-.037)	(-.037)	$\alpha$	(-.051)	(-.048)	(-.044)	(-.043)	(-.043)	(-.042)	(-.041)	(-.040)	(-.039)	(-.039)	(-.038)	(-.038)	(-.038)	(-.038)	(-.038)	(-.037)	(-.037)	(-.037)	(-.037)	$\alpha$	(-.051)	(-.048)	(-.044)	(-.043)	(-.043)	(-.042)	(-.041)	(-.040)	(-.039)	(-.039)	(-.038)	(-.038)	(-.038)	(-.038)	(-.038)	(-.037)	(-.037)	(-.037)	(-.037)	$\alpha$	(-.051)	(-.048)	(-.044)	(-.043)	(-.043)	(-.042)	(-.041)	(-.040)	(-.039)	(-.039)	(-.038)	(-.038)	(-.038)	(-.038)	(-.038)	(-.037)	(-.037)	(-.037)	(-.037)	$\alpha$	(-.051)	(-.048)	(-.044)	(-.043)	(-.043)	(-.042)	(-.041)	(-.040)	(-.039)	(-.039)	(-.038)	(-.038)	(-.038)	(-.038)	(-.038)	(-.037)	(-.037)	(-.037)	(-.037)	$\omega$	(.003)	(.004)	(.003)	(.001)	(.000)	(.002)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	$\omega$	(.003)	(.004)	(.003)	(.001)	(.000)	(.002)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	$\omega$	(.003)	(.004)	(.003)	(.001)	(.000)	(.002)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	$\omega$	(.003)	(.004)	(.003)	(.001)	(.000)	(.002)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	$\omega$	(.003)	(.004)	(.003)	(.001)	(.000)	(.002)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)







Table D.10  
 サンプルサイズ 50 における平均 CFI (上段) / RMSEA (下段)

閾値の均質性	モデル A			モデル B			モデル C		
	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2
2 カテゴリ									
均質	.977	.850	.793	.987	.984	.974	-	-	-
	.028	.092	.112	.025	.030	.042			
中程度に不均質	.974	.853	.790	.986	.982	.974	-	-	-
	.030	.090	.112	.026	.032	.042			
かなり不均質	.976	.838	.763	.987	.984	.977	-	-	-
	.024	.086	.107	.018	.022	.031			
極めて不均質	.564	.476	.419	.798	.872	.528	-	-	-
	.068	.071	.088	.059	.049	.082			
3 カテゴリ									
均質	.984	.829	.777	-	-	.975	-	-	-
	.025	.116	.136			.048			
中程度に不均質	.982	.844	.786	-	.981	.973	-	-	-
	.029	.113	.134		.035	.053			
かなり不均質	.981	.847	.795	-	-	.970	-	-	-
	.031	.112	.132			.054			
4 カテゴリ									
均質	-	-	-	-	-	-	-	-	-
中程度に不均質	-	-	-	-	-	-	-	-	-
かなり不均質	-	-	-	-	-	-	-	-	-
5 カテゴリ									
均質	-	-	-	-	-	-	-	-	-
中程度に不均質	-	-	-	-	-	-	-	-	-
かなり不均質	-	-	-	-	-	-	-	-	-
6 カテゴリ									
均質	-	-	-	-	-	-	-	-	-
中程度に不均質	-	-	-	-	-	-	-	-	-
かなり不均質	-	-	-	-	-	-	-	-	-

Table D.11  
 サンプルサイズ 100 における平均 CFI (上段) / RMSEA (下段)

閾値の均質性	モデル A			モデル B			モデル C		
	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2
2 カテゴリ									
均質	.989	.852	.795	.992	.987	.975	.992	.974	.954
	.018	.093	.110	.018	.026	.043	.013	.029	.040
中程度に不均質	.989	.845	.787	.993	.989	.975	.990	.973	.955
	.018	.093	.111	.018	.025	.042	.015	.028	.037
かなり不均質	.987	.845	.785	.992	.989	.977	.989	.973	.955
	.018	.090	.106	.018	.022	.038	.015	.027	.036
極めて不均質	.874	.760	.669	.904	.891	.873	-	.952	.923
	.022	.036	.044	.028	.029	.032	-	.008	.015
3 カテゴリ									
均質	.992	.837	.788	.996	.991	.974	-	-	.950
	.019	.115	.132	.015	.027	.055	-	-	.049
中程度に不均質	.992	.841	.788	.995	.990	.975	-	-	.948
	.017	.114	.132	.017	.028	.054	-	-	.050
かなり不均質	.992	.843	.791	.995	.991	.975	-	-	.947
	.017	.112	.130	.016	.027	.053	-	-	.050
4 カテゴリ									
均質	.994	.842	.796	.996	.990	.970	-	-	-
	.017	.124	.142	.017	.033	.066	-	-	-
中程度に不均質	.993	.839	.793	.996	.990	.972	-	-	-
	.019	.123	.140	.017	.033	.063	-	-	-
かなり不均質	.992	.841	.793	.995	.989	.972	-	-	-
	.020	.120	.137	.019	.033	.060	-	-	-
5 カテゴリ									
均質	.993	.840	.795	.996	.990	.970	-	-	-
	.021	.133	.151	.017	.035	.070	-	-	-
中程度に不均質	.993	.840	.793	.996	.990	.972	-	-	-
	.020	.130	.149	.017	.036	.066	-	-	-
かなり不均質	.991	.837	.790	.995	.989	.972	-	-	-
	.023	.127	.145	.021	.035	.063	-	-	-
6 カテゴリ									
均質	.993	.834	.792	.996	.990	.971	-	-	-
	.023	.148	.166	.021	.041	.076	-	-	-
中程度に不均質	.992	.837	.791	.996	.990	.971	-	-	-
	.024	.143	.163	.021	.040	.073	-	-	-
かなり不均質	.991	.835	.786	.995	.989	.971	-	-	-
	.024	.139	.158	.020	.037	.070	-	-	-

Table D.12  
 サンプルサイズ 200 における平均 CFI (上段) / RMSEA (下段)

閾値の均質性	モデル A			モデル B			モデル C		
	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2
2 カテゴリ									
均質	.995	.846	.791	.997	.992	.976	.996	.975	.953
	.012	.096	.112	.012	.022	.045	.009	.031	.042
中程度に不均質	.994	.849	.790	.996	.991	.976	.996	.974	.953
	.012	.093	.110	.012	.023	.043	.009	.030	.040
かなり不均質	.994	.848	.788	.996	.992	.978	.995	.977	.957
	.013	.090	.106	.012	.020	.040	.010	.028	.038
極めて不均質	.993	.866	.754	.994	.993	.988	1.00	.998	.984
	.003	.029	.040	.004	.005	.008	.000	.001	.006
3 カテゴリ									
均質	.997	.838	.789	.998	.991	.971	.997	.971	.943
	.011	.117	.133	.010	.030	.060	.008	.041	.055
中程度に不均質	.996	.841	.791	.998	.992	.972	.997	.969	.942
	.011	.115	.132	.012	.029	.059	.009	.041	.054
かなり不均質	.996	.842	.794	.998	.991	.972	.997	.972	.945
	.012	.113	.129	.011	.029	.058	.010	.039	.053
4 カテゴリ									
均質	.998	.837	.798	.998	.991	.968	.997	.965	.934
	.010	.127	.141	.011	.035	.070	.009	.047	.063
中程度に不均質	.997	.837	.796	.998	.991	.971	.998	.967	.938
	.011	.125	.140	.010	.033	.065	.008	.045	.059
かなり不均質	.996	.838	.789	.998	.991	.971	.997	.970	.942
	.012	.121	.138	.012	.032	.063	.009	.042	.057
5 カテゴリ									
均質	.997	.838	.798	.998	.990	.967	.997	.965	.932
	.012	.133	.149	.012	.039	.074	.010	.050	.067
中程度に不均質	.997	.839	.796	.998	.991	.970	.997	.965	.936
	.012	.129	.145	.011	.037	.069	.010	.048	.062
かなり不均質	.996	.839	.796	.998	.991	.971	.997	.970	.941
	.013	.125	.141	.012	.034	.064	.011	.043	.059
6 カテゴリ									
均質	.997	.836	.793	.998	.990	.967	.997	.964	.931
	.013	.140	.158	.012	.042	.078	.011	.053	.071
中程度に不均質	.997	.839	.795	.998	.990	.969	.997	.963	.932
	.014	.136	.153	.013	.040	.073	.011	.052	.068
かなり不均質	.996	.836	.794	.998	.990	.969	.996	.966	.936
	.014	.133	.149	.014	.038	.071	.013	.048	.065

Table D.13  
 サンプルサイズ 500 における平均 CFI (上段) / RMSEA (下段)

閾値の均質性	モデル A			モデル B			モデル C		
	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2
2 カテゴリ									
均質	.998	.843	.789	.999	.992	.974	.998	.974	.948
	.007	.097	.112	.007	.025	.048	.006	.033	.046
中程度に不均質	.998	.848	.792	.999	.992	.977	.998	.973	.950
	.006	.094	.109	.008	.024	.045	.006	.033	.044
かなり不均質	.998	.848	.789	.998	.994	.977	.998	.976	.954
	.008	.091	.107	.007	.021	.043	.006	.030	.041
極めて不均質	.999	.866	.757	1.00	.999	.995	1.00	.997	.969
	.001	.036	.049	.001	.001	.007	.000	.003	.015
3 カテゴリ									
均質	.999	.837	.789	.999	.991	.969	.999	.968	.936
	.007	.119	.134	.007	.033	.063	.006	.044	.060
中程度に不均質	.999	.837	.790	.999	.992	.971	.999	.967	.937
	.007	.118	.133	.006	.032	.061	.005	.043	.059
かなり不均質	.999	.838	.792	.999	.991	.971	.999	.970	.939
	.007	.115	.130	.007	.032	.060	.005	.042	.057
4 カテゴリ									
均質	.999	.835	.795	.999	.990	.966	.999	.962	.926
	.007	.128	.142	.007	.039	.072	.006	.050	.068
中程度に不均質	.999	.840	.799	.999	.990	.968	.999	.962	.930
	.007	.123	.138	.008	.037	.068	.006	.048	.064
かなり不均質	.999	.839	.795	.999	.991	.970	.999	.968	.936
	.007	.121	.136	.007	.034	.064	.006	.044	.060
5 カテゴリ									
均質	.999	.835	.796	.999	.990	.965	.999	.962	.923
	.007	.134	.149	.006	.042	.077	.006	.052	.072
中程度に不均質	.999	.837	.797	.999	.990	.968	.999	.961	.928
	.007	.129	.144	.007	.039	.071	.006	.051	.067
かなり不均質	.999	.838	.795	.999	.991	.970	.999	.968	.936
	.007	.124	.139	.007	.035	.066	.006	.045	.062
6 カテゴリ									
均質	.999	.832	.794	.999	.989	.964	.999	.960	.921
	.008	.140	.154	.007	.044	.080	.006	.055	.075
中程度に不均質	.999	.838	.799	.999	.989	.966	.999	.959	.923
	.008	.134	.149	.007	.042	.076	.006	.054	.072
かなり不均質	.999	.837	.796	.999	.991	.967	.999	.964	.930
	.008	.130	.145	.007	.039	.073	.006	.050	.068

Table D.14

サンプルサイズ 1000 における平均 CFI (上段) / RMSEA (下段)

閾値の均質性	モデル A			モデル B			モデル C		
	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2	正しく特定	誤特定 1	誤特定 2
2 カテゴリ									
均質	.999	.844	.790	.999	.992	.973	.999	.972	.945
	.005	.097	.112	.005	.026	.049	.004	.035	.048
中程度に不均質	.999	.845	.789	.999	.993	.975	.999	.972	.948
	.005	.095	.111	.005	.025	.047	.004	.034	.045
かなり不均質	.999	.844	.785	.999	.993	.976	.999	.977	.953
	.005	.093	.108	.005	.023	.044	.004	.031	.042
極めて不均質	.999	.852	.749	1.00	.999	.991	1.00	.991	.959
	.001	.044	.057	.001	.002	.013	.001	.010	.022
3 カテゴリ									
均質	.999	.837	.791	1.00	.991	.969	.999	.966	.932
	.005	.119	.134	.005	.034	.064	.004	.046	.062
中程度に不均質	.999	.837	.791	1.00	.991	.970	.999	.965	.933
	.004	.118	.133	.005	.034	.062	.004	.045	.061
かなり不均質	.999	.838	.791	.999	.991	.970	.999	.968	.936
	.005	.116	.131	.005	.033	.061	.004	.043	.059
4 カテゴリ									
均質	.999	.835	.795	1.00	.990	.964	1.00	.961	.922
	.005	.129	.143	.005	.040	.074	.004	.052	.070
中程度に不均質	.999	.839	.796	1.00	.990	.967	.999	.960	.925
	.005	.125	.139	.004	.038	.070	.004	.050	.067
かなり不均質	.999	.838	.794	1.00	.991	.968	.999	.966	.932
	.005	.121	.136	.005	.036	.066	.004	.046	.063
5 カテゴリ									
均質	.999	.832	.795	1.00	.989	.964	1.00	.959	.919
	.005	.135	.149	.005	.042	.078	.004	.055	.074
中程度に不均質	.999	.838	.798	1.00	.990	.967	.999	.958	.924
	.005	.129	.144	.005	.040	.072	.004	.053	.069
かなり不均質	.999	.837	.794	1.00	.991	.968	.999	.966	.932
	.004	.124	.139	.005	.037	.067	.004	.047	.064
6 カテゴリ									
均質	.999	.834	.797	1.00	.989	.963	.999	.958	.917
	.005	.138	.152	.005	.044	.081	.004	.057	.077
中程度に不均質	1.00	.835	.796	1.00	.989	.965	1.00	.958	.920
	.004	.135	.149	.005	.043	.077	.004	.055	.073
かなり不均質	.999	.837	.796	1.00	.990	.966	.999	.962	.925
	.005	.130	.145	.005	.040	.074	.004	.051	.070

# 初出一覧

## 第 1 章：序論

本論文のための書き下ろし。

## 第 2 章：順序カテゴリデータから計算される相関係数の数理

以下の論文の一部をもとに加筆修正した。

- Onoshima, T., Shiina, K., Ueda, T., & Kubo, S. (2019). Decline of Pearson's  $r$  with categorization of variables: A large-scale simulation. *Behaviormetrika*, 46(2), 380 – 399. <https://doi.org/10.1007/s41237-019-00089-1>

## 第 3 章：カテゴリ化によって生じる相関係数のバイアスの評価

上記論文 (Onoshima et al., 2019) に基づいている。また、結果の一部は以下の学会で発表された。

- 小野島昂洋・椎名乾平・上田卓司・久保沙織 (2019) . 連続変数の離散化に伴う相関係数の低下——離散化の際の閾値を考慮した大規模シミュレーション——. 日本心理学会第 83 回大会.



## 第4章：カテゴリ化によって生じる相関係数のバイアスへの対処法の評価

以下の論文を加筆修正した。

- 小野島昂洋・椎名乾平（2019）． カテゴリカルな教育データと相関係数に関する留意点． コンピュータ&エデュケーション, 47, 25-30.  
<https://doi.org/10.14949/konpyutariyoukyouiku.47.25>

## 第5章：カテゴリ化が信頼性係数に与える影響

以下の論文を加筆修正した。

- 小野島昂洋・椎名乾平（2021）． 順序カテゴリデータへの確認的因子分析に基づく信頼性係数の評価——モデルが正しく特定された場合と誤特定された場合の比較——． 教育心理学研究, 69(3), 281-296.

また、結果の一部は以下の学会で発表された。

- 小野島昂洋・椎名乾平（2020）． 変数のカテゴリ化によって生じる相関係数のバイアスがクロンバックの  $\alpha$  へと及ぼす影響． 日本教育心理学会第62回総会.

## 第6章：結論と展望

本論文のための書き下ろし。