

論文 / 著書情報
Article / Book Information

題目(和文)	多次元項目反応理論モデルにおける共通受検者を利用した等化法の提案
Title(English)	
著者(和文)	沖嘉訓
Author(English)	Yoshinori Oki
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第10572号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:前川 眞一,林 直亨,中山 実,室田 真男,松田 稔樹
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第10572号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

多次元項目反応理論モデルにおける
共通受検者を利用した等化法の提案

沖 嘉訓

目次

第1章	研究の背景と目的	3
第2章	IRT のモデル	8
2.1	1次元のIRTモデル	9
2.1.1	1パラメータ・ロジスティックモデル	9
2.1.2	2パラメータ・ロジスティックモデル	11
2.1.3	3パラメータ・ロジスティックモデル	12
2.1.4	段階反応モデル	13
2.2	多次元IRTモデル	16
2.2.1	2パラメータ・ロジスティック補償型モデル	16
2.2.2	2パラメータ・ロジスティック部分補償型モデル	19
2.2.3	多次元段階反応モデル	19
2.2.4	多次元IRTと因子分析モデルの関係	23
第3章	多次元IRTの等化の先行研究	26
3.1	多次元IRTのパラメータの不定性	27
3.2	多次元IRTの等化	28
3.3	多次元IRTの等化法(先行研究)	29
3.3.1	TCF法	29
3.3.2	ICF法	29
3.3.3	Li & Lissitzの方法	30
3.3.4	Minの方法	31
3.3.5	NOP法	31
3.3.6	Hirschの方法	32
第4章	共通受検者を利用した多次元IRTの等化法	36
4.1	問題	37
4.2	提案手法	37
4.2.1	受検者特性値の一致度を高める方法	38
4.2.2	相関係数を一致させる方法	39
4.3	計算機による実験	41
4.3.1	次元間の相関に関する実験	41
4.3.2	項目数に関する実験	46

4.4	まとめと課題	49
第5章	共通受検者と共通項目を同時に利用した 多次元IRTの等化法	50
5.1	問題	51
5.2	共通項目と共通受検者の両方を考慮した等化法	51
5.2.1	共通項目に注目した重み付きプロクラステス回転による方法	52
5.2.2	共通項目と共通受検者の両方を考慮した重み付きプロクラステス回転による方法	53
5.3	計算機による実験	56
5.3.1	次元間の相関係数に関する実験	57
5.3.2	次元の数に関する実験	66
5.4	まとめと課題	72
第6章	実データを用いた提案手法の評価	74
6.1	分析手順	75
6.2	結果	76
第7章	全体のまとめ	94
7.1	本研究の目的	95
7.2	本研究の結果と意義	96
7.3	今後の課題	99

第1章 研究の背景と目的

テストは私達の生活に浸透している。学校では日常的にクラス単位でテストが実施されている。また、学校をまたいだ統一的な学力テストが行われることもある。あるいは、OECD 学習到達度調査 (PISA) や国際数学・理科教育動向調査 (TIMSS) のように国をまたいで実施される国際的なテストも存在する。学校の外でも、TOEIC や TOEFL, 各種の資格試験のように幅広い受検者を対象として能力を診断するテストが広く普及している。

日本テスト学会は、テストの作成、実施、採点、結果の利用、管理など一連のテストの開発、実施過程でのあるべき姿と規準を「テスト・スタンダード」としてまとめている (日本テスト学会 (編), 2007)。そこでは対象となるテストを「ある目的を遂行するために、原則としてある程度の期間安定して持続する人間の特性を測定するための用具」と定義している。2015 年から一定規模以上の企業についてストレス診断の実施が義務化されたが、上記の定義にはこのような学力以外の特性を診断するシステムもテストの一種として含まれる。

さて、学校で実施されるクラス単位のテストの場合、1 つのテストの受検者は数十人から数百人程度であることが多いと思われるが、テストの中には年間の受検者数が数万人から数百万人に及ぶような大規模なものが存在する。また、たとえば TOEIC のように毎年数回実施されるテストもある。これらのテストでは、同一の内容を測定する問題項目の異なる複数のバリエーション (以下テストフォームと呼ぶ) が用意されることがある。たとえば毎年複数回実施される英語のテストについて、同じ受検者が繰り返しテストを受検することを許容する場合、テストフォームが 1 種類のみでは、以前受検した問題が次回も出題されることになり、ただしく個人の能力を診断することができない。大規模なテストの場合にも、なんらかの都合で受検者間で受検時期にずれが生じた場合、問題内容の漏洩が心配される。複数のテストフォームを用意することで、これらの影響を一定程度抑制することが可能となる。

一方で、複数のテストフォームを準備した場合には、フォーム間でのスコアの比較可能性という課題が生じる。テストのスコアリングでは、正解率を 100 点満点換算した値を個人のスコアとして利用するような方法がよく利用されるが、この方法ではテストフォーム毎の問題項目の難易度の差がスコアに影響を及ぼすため、公平性やスコアの解釈という意味で問題がある。これを解決する技術として、等化と呼ばれる手続きが存在する。等化は、複数のテストフォームのスコアを、同一の尺度上で比較可能な形に変換する操作である。図 (1.1) は、大規模テストの開発の流れの例を示したものである。等化が必要となる大規模テストでは、本格的なテストの実施に先立って、問題項目を仮のテストフォームに編集し、プリテストと呼ばれる項目の選別と等化を目的としたデータ収集のための予備的なテストが実施されることが多い。テストを測定の用具として考えることは、テスト自体の性質を評価するという視点につながるが、上記のような開発プロセスは、テストスコアの比較可能性や広い意味

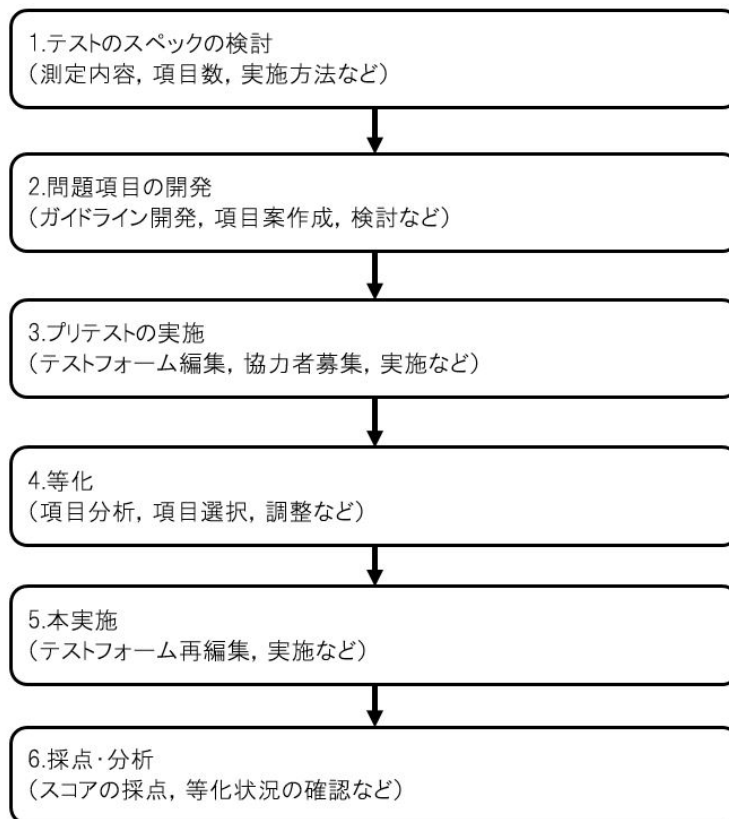


図 1.1: 大規模テストの開発フローの例

での妥当性 (e.g. 村山, 2012) を確保するための工夫として有効なものである。

さて, 学力検査やパーソナリティ検査などのテストの分析および採点に利用される統計モデルとして項目反応理論 (項目応答理論, ItemResponseTheory; Lord,1980; 芝,1991; 池田,2001; 村木,2011; 加藤・山田・川端,2014 :以下 IRT) がある。IRT の特徴は, 問題項目に対する回答を受検者の潜在的な特性と問題項目の特性による確率的な関数として表現し, これらの特性を分離して評価する点にある。テストの分析に IRT を利用することの大きな利点として, 等化に対する自由度が高いことが挙げられる。本邦では近年, 大学入試改革に関する議論が進められているが, この流れの中でも特に高等学校基礎学力テスト (仮称) において IRT の活用が検討されている。この背景にも, 複数のテストの結果の比較可能性を確保することに対する要請がある (高大接続システム改革プロジェクトチーム, 2016)。

さて, 分析対象とする問題項目の形式や, 受検者の特性を表す潜在変数と回答データとの関連性をどのように捉えるかということに対応して, IRT には複数の下位モデルが開発されている。このうち, 多次元項目反応モデル (Multidi-

mensional Item Response Theory Model; Reckase,1985, Reckase & McKinley,1991, Reckase,2009 :以下多次元 IRT) は項目に対する回答の背後に複数の能力が同時に関与することを想定し, 受検者を多次元の潜在空間における特性値ベクトルとして位置づけるモデルである.

多次元 IRT においても等化はやはり重要な課題であり, 各種の等化法に関する研究が進められている (e.g. Simon,2008; Jonathan,2014)¹. 多次元 IRT の主な等化法には, 1次元の IRT と同様に, テストフォーム間で共通する問題項目を利用する共通項目による方法と, 共通する受検者を利用する共通受検者による方法がある. このうち, 共通項目による等化法としては, 1次元の IRT の場合の等化法を多次元に拡張した複数の方法 (Oshima, Davey & Lee, 2000) が提案されているほか, 多次元 IRT に固有の方法として Li & Lissitz の方法 (Li & Lissitz,2000), Min の方法 (Min,2003) などが開発されている. もう一方の共通受検者による等化法に関する研究としては Hirsch によるもの (Hirsch,1989) を挙げることができる. しかし, 共通受検者による多次元 IRT の等化法の研究は, その他にほとんど見あたらず, 共通項目に関する研究と比較して非常に数が少ないのが現状である. また, Hirsch の方法は 2つのテストフォームの共通受検者について, 基底を揃える変換を行った後に直交プロクラステス回転を行うことで等化を実現するものだが, 後で詳しく見るように, この方法では等化後の受検者特性値について次元間の相関構造が等化先と一致しないなどいくつかの課題を指摘することができる.

一方で, テストの実務において共通受検者による等化が求められる場面は少なからず存在するため, 多次元 IRT の活用場面の拡大のためには, 特に共通受検者による等化法の充実が求められるものといえる. 本研究は多次元 IRT における共通受検者による等化の問題にフォーカスし, 先行研究の課題を克服したいくつかの等化法を新たに提案することを目的とする.

本研究の構成は, 以下の通りである. まず, 本章につづく第 2 章において, 1次元の IRT のモデルと多次元 IRT のモデルを確認する. つぎに, 第 3 章において多次元 IRT の等化を定式化し, 先行研究で示された等化法について概要を確認する. この際, 受検者特性値による等化法である Hirsch の方法については, 特に詳しくその特徴を調べる. 第 4 章では, 受検者特性値による新たな等化法として, 受検者特性値に注目した斜交プロクラステス回転法と, 同じく受検者特性値に注目した重み付きプロクラステス回転法の 2つの等化法を提案し, 計算機実験によってその特性を既存の等化法と比較する. 第 5 章では, 第 4 章の実験結果を考慮した上で, 共通受検者による情報に共通項目の情報も併せて利用したハイブリッドな等化法を提案し, 計算機実験によって特性を確認する. 第 6 章では OECD 学習到達度調査 PISA の実際の回答データに対して受検者特性値に注目した重み付きプロクラステス回転法を適用し, 等化の結果について確認する. 第 7 章で, 全体の内容についてまとめ, 今後の研究の展

¹ただし, Simon(2008) は多次元 IRT の等化法の研究が十分ではないことを指摘している.

開可能性について述べる. なお, 各章の関係性は次の図 (1.2) の通りである.

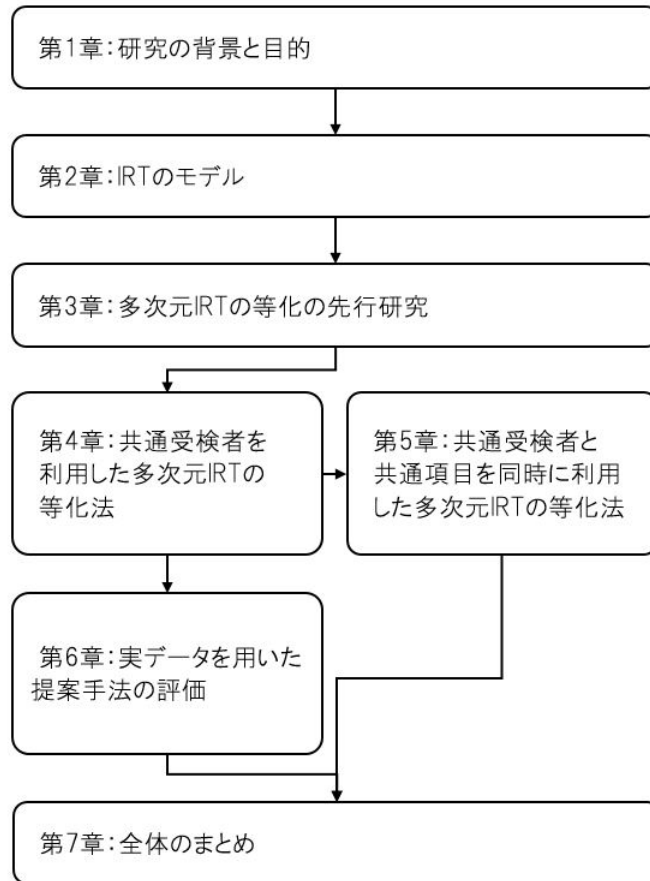


図 1.2: 各章の関係

第2章 IRTのモデル

2.1 1次元のIRTモデル

IRTには、受検者特性と項目反応に想定される関係性に対応した多数の下位モデルが提案されている。(e.g. van der Linden & Hambleton, 1997; Reise & Revicki, 2014). それらのモデルのほとんどは、項目に対する回答に単次元の受検者特性が関係することを仮定する1次元モデルである。ここではまず、1次元のIRTの基本的なモデルについて概観し、その後には本研究の対象である多次元のモデルについて確認する。

以降で登場する数式の表記の規則はつぎの通りである。まず、 α , θ のような、ボールド体小文字のアルファベットおよびギリシャ文字はベクトルを表す。ベクトルは縦ベクトルとする。ボールド体でない通常の小文字はスカラーである。また、 A , Θ のような大文字は行列を表す。添え字については、各数式が初出の場合に、その意味を明記するようにした。¹

2.1.1 1パラメータ・ロジスティックモデル

1パラメータ・ロジスティックモデルはIRTの最も基本的なモデルであり、受検者の特性と問題項目に対する正誤に以下の確率的な関係を想定する。

$$P(x_j = 1|\theta, b_j) = \frac{\exp(\theta - b_j)}{1 + \exp(\theta - b_j)} \quad (2.1)$$

上式の x_j は項目 j に対する2値の回答で、 $x_j = 1$ は正答を、 $x_j = 0$ は誤答を表す。また、 θ は受検者の能力特性値を表す潜在変数である。図2.1は、 b_j の値の異なる2つの項目について、1パラメータ・ロジスティックモデルにおける θ と $P(x_j|\theta, b_j)$ の関係性を、横軸を受検者特性値 θ 、縦軸を正答確率 $P(x_j|\theta, b_j)$ としてグラフに表したものである。この図のように、能力特性値と正答確率の関係を表した曲線は、項目特性曲線 (Item Characteristic Curve) と呼ばれる。

なお、正答と誤答が背反であることから、

$$P(x_j = 0|\theta, \xi_j) = 1 - P(x_j = 1|\theta, \xi_j) \quad (2.2)$$

である。ただし、 ξ は項目に関するパラメータをまとめて表現した変数であり、この関係は以下で確認する正誤2値に対応したIRTの諸モデルに共通である。

ところでIRTは、テストのデータを分析するために広く用いられる確率的なモデルであるが、分析の対象として取り上げられるテストはいわゆる学力検査に限られるものではなく、パーソナリティに関連する心理尺度の分析にIRTが利用されることも多い。また組織の特性に対する調査をIRTの枠組みで分析した事例も存在する(渡辺・野口, 1999)。データが学力テスト以

¹ベクトル、スカラーの両方の可能性がある変数についてはベクトルとして、同じく行列、ベクトルの両方の可能性がある変数については行列として表記する。

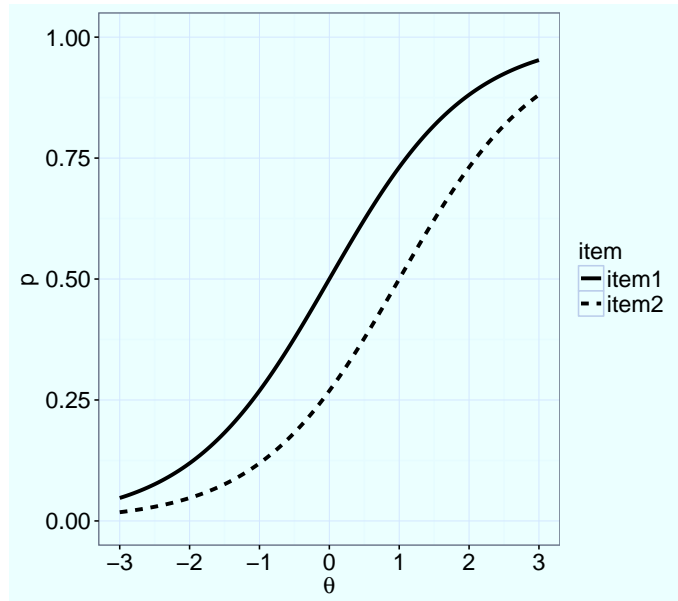


図 2.1: 1 パラメータ・ロジスティックモデルの項目特性曲線 (item1: $b = 0$; item2: $b = 1$ の場合)

外の場合には、正答や誤答、あるいは受検者などの表現がそぐわない面もあるが、本研究では便宜的に学力検査を想定した用語を用いることとする。

表 2.1: 1 パラメータ・ロジスティックモデルの項目パラメータ

	b
item1	0.0
item2	1.0

さて、図 2.1 をみると、受検者特性値 θ がどの位置であっても、項目パラメータ b の値が $b = 0$ の場合のほうが、 $b = 1$ の場合よりも常に縦軸の正答確率が高くなっていることがわかる。このことは、どの受検者にとっても $b = 0$ の問題のほうが $b = 1$ の問題よりも、正答するのが容易な問題項目であることを表すものと解釈できる。

1 パラメータ・ロジスティックモデルの特徴は、項目パラメータ b_j によって個々の問題項目の難易度の違いを表現することで、受検者の特性とは分離する形で問題項目の特性を取り扱う点にある。このように問題項目と受検者の特性の関係を確率的なモデルとして表現することで、問題項目の特性を受検者の特性から切り離して検討可能とする点は、項目反応理論のモデルに共通したコンセプトである²。

²1 パラメータ・ロジスティックモデルは、ラッシュ・モデルと呼ばれるモデルと数学的には

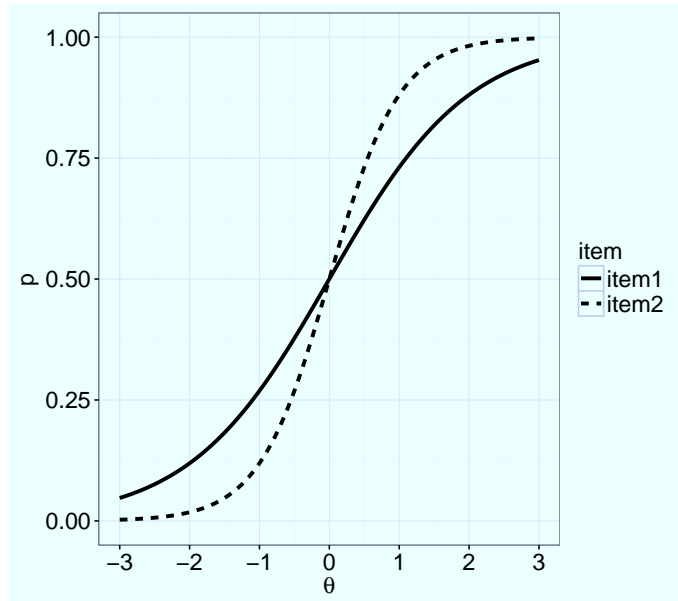


図 2.2: 2パラメータ・ロジスティックモデルの項目特性曲線 (item1: $a = 1, b = 0$; item2: $a = 2, b = 0$ の場合)

2.1.2 2パラメータ・ロジスティックモデル

1パラメータ・ロジスティックモデルでは、項目の特徴を表すパラメータが難易度に関連する b_j の1種類であった。2パラメータ・ロジスティックモデルは、困難度 b_j に加えて識別力 a_j を用いて項目の特徴を表現するモデルである。2パラメータ・ロジスティックモデルは、以下の式で表現される。

$$P(x_j = 1|\theta, a_j, b_j) = \frac{\exp(a_j(\theta - b_j))}{1 + \exp(a_j(\theta - b_j))} \quad (2.3)$$

1パラメータ・ロジスティックモデルと同様、上式の x_j は項目 j に対する2値の回答データを表す。また、 θ は受検者の能力特性を表す潜在変数である。

表 2.2: 2パラメータ・ロジスティックモデルの項目パラメータ

	a	b
item1	1.0	0.0
item2	2.0	0.0

同一のモデルである。ただし、両モデルは歴史的には別々に開発されたものであり、背景とする考え方には違いがある。具体的にはラッシュ・モデルでは、測定の不変性すなわち問題項目に依存しない受検者特性値の測定、あるいは受検者群に依存しない項目特性の把握という点により力点が置かれる。詳細は、静 (2007), 村木 (2011), 加藤他 (2014) 等を参照。

図(2.2)は、 b パラメータを0に固定し、 a パラメータを1.0と2.0とした2項目について、項目特性曲線を比較したものである。図から $a = 2.0$ の項目 (item2) のほうが、 $a = 1.0$ の項目 (item1) よりも $\theta = 0$ 付近での傾きが急峻になっていることがわかる。

2パラメータ・ロジスティックモデルにおける a パラメータは $\theta = b_j$ における項目特性曲線の傾きに対応した変数であり項目識別力パラメータと呼ばれる。通常 $a > 0$ である。 a パラメータの値が大きくと項目特性曲線の傾きが急である場合、 a の値が小さい場合に比較して、 $\theta < b_j$ の領域での正答確率が小さくなり、 $\theta > b_j$ の領域での正答確率がより大きくなる。テストの問題項目について、能力が一定以上のレベルであれば正答確率が高くなり、一定レベルを下回れば正答確率が低くなるとするのは自然な想定であると思われるが、その程度については問題項目の特性によって異なる可能性がある。2パラメータ・ロジスティックモデルは受検者特性値と正答確率のこのような関係性の程度を a パラメータを導入することで表現したものと解釈できる。

表 2.3: 2パラメータ・ロジスティックモデルの項目パラメータ

	a	b
item1	2.0	0.0
item2	2.0	1.0

図(2.3)は、 a パラメータを2.0に固定して、 b の値を0.0と1.0とした2項目の項目特性曲線を示したものである。 a パラメータの値が同じ場合には2つの項目特性曲線の関係は1パラメータ・ロジスティックモデルと類似しており、2パラメータ・ロジスティックモデルにおいても b パラメータは項目の困難度を表現していることが理解できる。

2.1.3 3パラメータ・ロジスティックモデル

3パラメータ・ロジスティックモデルは、2パラメータ・ロジスティックモデルにさらに項目パラメータ c_j を加えたモデルで、次式で表される。

$$P(x_j = 1|\theta, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{\exp(a_j(\theta - b_j))}{1 + \exp(a_j(\theta - b_j))} \quad (2.4)$$

図(2.4)は、各項目パラメータを表(2.4)の値とし、 c パラメータの値の異なる2つの項目について、項目特性曲線を描画したものである。

グラフから、 c パラメータの値が大きい場合には、 θ が非常に低い値の場合の正答確率が高くなっており、 c パラメータの大きさが能力レベルが非常に低い受検者の当該項目に正答する確率に対応していることがわかる。3パラメータ・ロジスティックモデルは、能力の面では本来正答が難しい受検者が、当該

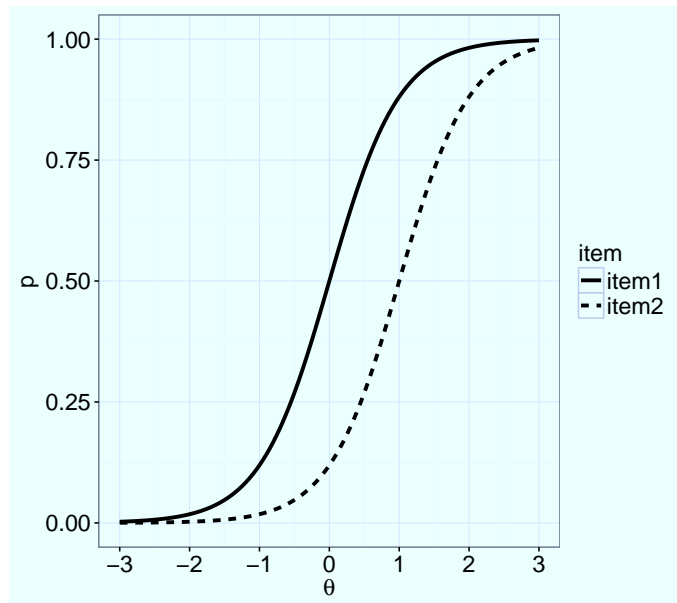


図 2.3: 2パラメータ・ロジスティックモデルの項目特性曲線 (item1: $a = 2, b = 0$; item2: $a = 2, b = 1$ の場合の例)

項目に偶然正答する現象を考慮したモデルとして理解できる³。

表 2.4: 3パラメータ・ロジスティック単次元モデルの項目パラメータ

	a	b	c
item1	2.0	0.0	0.1
item2	2.0	0.0	0.2

2.1.4 段階反応モデル

これまで、正誤2値データに対応する1次元のIRTのモデルを見てきたが、たとえば性格検査では、選択肢が「あてはまる」、「ややあてはまる」、「あまりあてはまらない」のような順序をもった段階から成るケースがよく見られる。学力テストにおいても、各問題項目に対する判定が³、正答、誤答の2値ではなく、「完全に正答」、「ほぼ正答」、「一部正答」、「完全に誤答」のように中間的な判定を許容するケースが存在する。

このように、回答データが多値の場合に対応したIRTのモデルもいくつか開発されている。ここでは多値型の1次元のIRTのモデルのうち、段階反応

³ただし、 c パラメータの解釈や取り扱いについては、様々な意見が存在する。詳細は例えば村木(2009)を確認のこと。

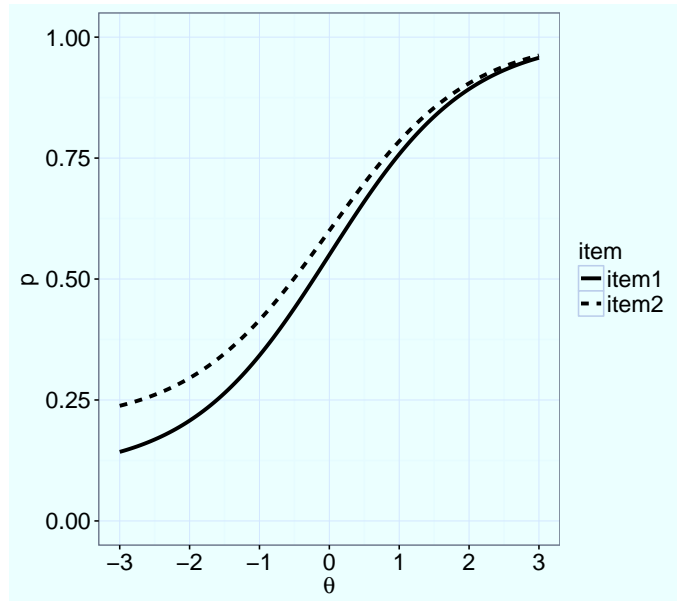


図 2.4: 3 パラメータ・ロジスティック単次元モデルの項目特性曲線 ($a = 2, b = 0, c = 0.1$ と $a = 2, b = 0, c = 0.2$ の場合の例)

モデル (Graded Response Model; Samejima, 1969) について確認する。

いま, 項目 j に対する反応が, $K_j + 1$ 個のカテゴリから成るものとする. 上記の学力テストの例で, 判定レベルが「3:完全に正答」, 「2:ほぼ正答」, 「1:一部正答」, 「0:完全に誤答」のように 4 つの場合には $K_j = 3$ である. 段階反応モデルでは, 受検者特性値を θ とし, 項目 j において, 受検者の回答がカテゴリ k ($k = 0, \dots, K_j$) と判定される確率を,

$$P_{jk}(\theta) = P_{jk}^+(\theta) - P_{j,k+1}^+(\theta) \quad (2.5)$$

であると想定する. 上式の $P_{jk}^+(\theta)$ は, 項目 j において k 以上の段階が選ばれる確率を表す. なお,

$$P_{j0}^+(\theta) = 1, \quad P_{j,K_j+1}^+(\theta) = 0 \quad (2.6)$$

とする.

各 k ($k = 1, \dots, K_j$) について $P_{j,k}^+(\theta)$ に前述したものと同様の 2 パラメータ・ロジスティックモデル

$$P_{j,k}^+ = \frac{\exp(a_j(\theta - b_{jk}))}{1 + \exp(a_j(\theta - b_{jk}))} \quad (2.7)$$

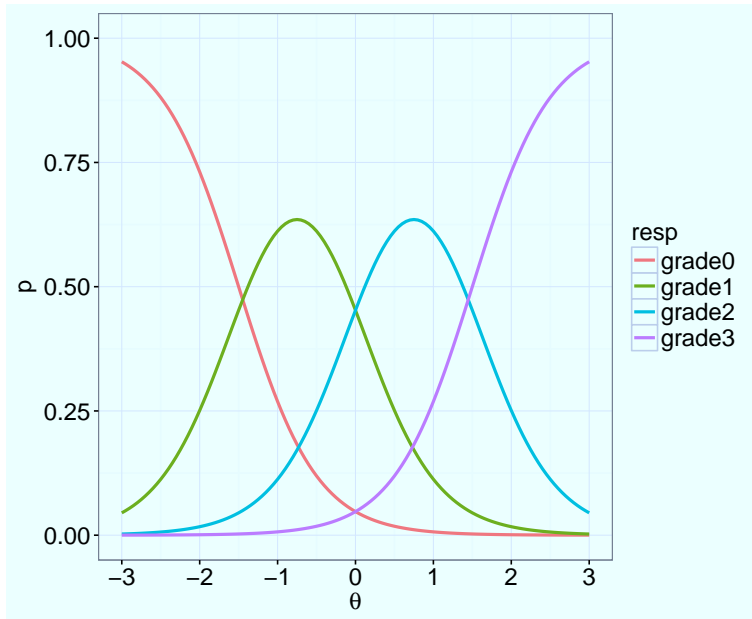


図 2.5: 段階反応モデルの項目特性曲線の例

を想定して、段階反応モデルの項目特性曲線⁴を描いたのが図 2.5 である。図に対応する各パラメータの値については、表 (2.5) にまとめる。

表 2.5: 段階反応モデルの項目パラメータ

a	b_1	b_2	b_3
2.0	-1.5	0.0	1.5

ところで、たとえば 2 パラメータ・ロジスティックモデルでは、正答確率が 0.5 となる能力特性値の位置については、当該項目の困難度パラメータ b の値に対応して決まる性質がある。一方、段階反応モデルでは段階 k が 1 から $K - 1$ の場合には、各段階の特性曲線の位置には 2 つの b パラメータが関係する。解釈の便利のために、段階反応モデルの場合の各段階の位置母数 b_{jk}^* をつぎのように定める。

$$b_{j0}^* = b_{j1} \quad , \quad (2.8)$$

$$b_{jk}^* = \frac{b_{jk-1} + b_{jk}}{2} \quad , 1 \leq k \leq K - 1 \quad (2.9)$$

$$b_{jK}^* = b_{jK} \quad . \quad (2.10)$$

⁴段階反応モデルの場合の能力特性値と各段階の確率の対応を表した曲線は、カテゴリ確率曲線と呼ばれる (加藤 他, 2014)。

段階反応モデルは、識別力パラメータ a を項目内で共通した値とし、困難度パラメータ b はカテゴリ毎に異なる値を取ることを許容するモデルである。その他の多値型の IRT のモデルには、名義反応モデル (Nominal Response Model; Bock, 1972), 部分採点モデル (Partial Credit Model; Masters, 1982), 一般化部分採点モデル (Generalized Partial Credit Model; Muraki, 1992) などが存在する。

2.2 多次元 IRT モデル

これまでテストに対応する受検者の能力が1次元の場合を想定した IRT のモデルを確認してきたが, Ackerman, Gierl & Walker (2005) や Yao & Boughton (2009) は, 多くのテストの回答の背後に多次元性が内在していることを指摘している。たとえば学力テストにおいて, 数学の問題の一部に文章題が含まれており, 正答のために計算能力に加えて文章理解能力も必要とされるような場合には, 多次元の能力特性を想定するのが自然である (孫,1997)。また, 性格検査において, 測定対象の2つの性格特性に対応した短文を対にしてどちらが当てはまるかを問う形式や, 複数の性格特性について自分にあてはまる順序を問う形式が利用されることがあるが, この場合は明らかに回答に複数の特性が同時に関与しており, 多次元 IRT を利用することが妥当であると考えられる (e.g. Brown & Maydeu-Olivares,2010; Brown & Maydeu-Olivares,2012)。

多次元項目反応モデル (Multidimensional Item Response Theory Model; Reckase,1985; Reckase & McKinley,1991; Reckase,2009 :以下, 多次元 IRT) は項目に対する回答の背後に複数の能力が同時に関与することを想定し, 受検者を多次元の潜在空間における特性値ベクトルとして位置づけるモデルである。回答データの形式や能力特性の各次元と項目への反応の関係をどのように捉えるかということに対応して, 多次元 IRT にもいくつかの下位モデルが開発されている。

ここでは代表的なモデルとして, 2パラメータ・ロジスティック補償型モデル (Two-Parameter Logistic Compensatory Model, Reckase,2009), 2パラメータ・ロジスティック部分補償型モデル (Two-Parameter Logistic Partial-compensatory Model, Reckase,2009), 多次元段階反応モデル (Multidimensional Graded Response Model, Muraki & Carlson,1995) の3つについて確認する。

2.2.1 2パラメータ・ロジスティック補償型モデル

2パラメータ・ロジスティック補償型モデルは, 次の式で表される。

$$P(x_j = 1|\boldsymbol{\theta}, \mathbf{a}_j, d_j) = \frac{\exp(\mathbf{a}_j^T \boldsymbol{\theta} + d_j)}{1 + \exp(\mathbf{a}_j^T \boldsymbol{\theta} + d_j)} . \quad (2.11)$$

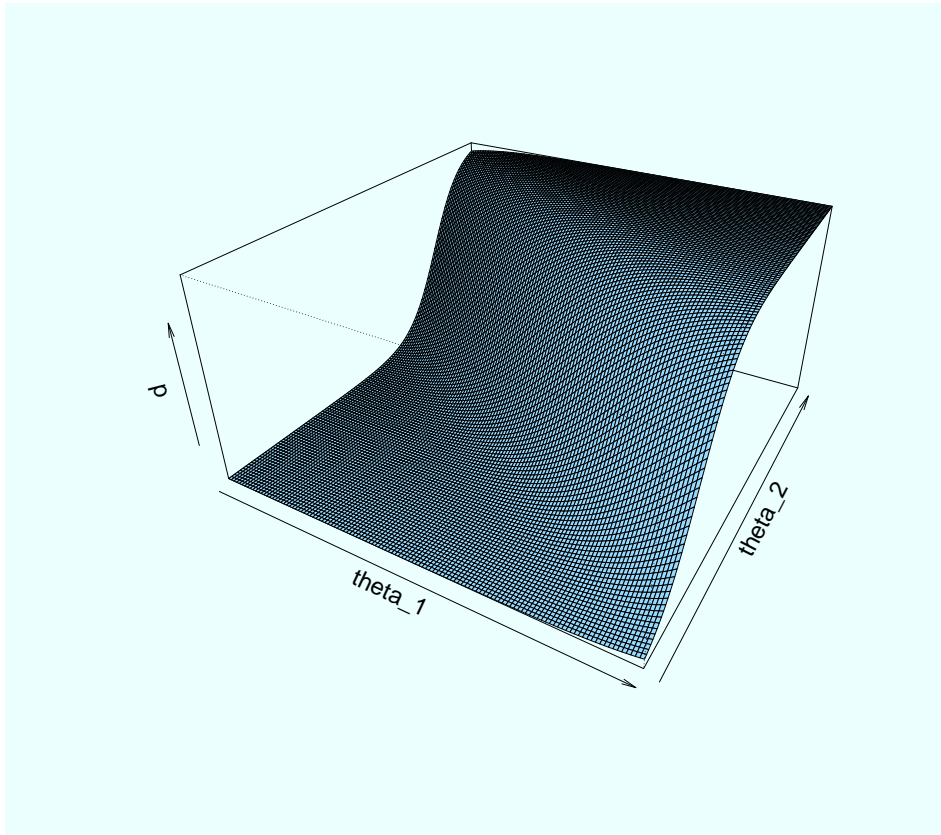


図 2.6: 2 パラメータ・ロジスティック補償型モデルの項目特性曲面の例

上式で x_j は項目 j について受検者が正答した場合に 1, 誤答の場合に 0 をとる 2 値の回答データである。また \mathbf{a}_j は項目 j の各次元の識別力を縦に並べたベクトル, d_j は項目 j の困難度に関連するパラメータ, $\boldsymbol{\theta}$ は受検者の各次元の能力特性値を縦に並べた能力特性値を表すベクトルである。能力特性値が多次元の場合を考慮するため, 受検者特性値 $\boldsymbol{\theta}$ はベクトルとなっている。これに対応して \mathbf{a}_j もベクトルの形で取り扱う。一方, 困難度に対応する d_j は当該項目の全般的な特徴としてスカラーで表現される。図 (2.6) および図 (2.7) は, 受検者特性の各次元と正答確率の関係を図示したものである。1 次元の場合の項目特性曲線に対して, このような曲面は項目特性曲面 (Item Characteristic Surface; ICS) と呼ばれる。補償型のモデルでは, 受検者特性の各次元の値と正答確率が加法的な関係となっており, ある次元において受検者特性値が低い場合であっても, 別の次元の能力特性値が高ければ, 正答確率はそれほど低くならない。項目特性曲面からも, この特徴が確認できる。

ここで, 1 次元の 2 パラメータ・ロジスティックモデルと多次元 IRT の 2 パラメータ・ロジスティック補償型モデルの関係について調べておく。前に

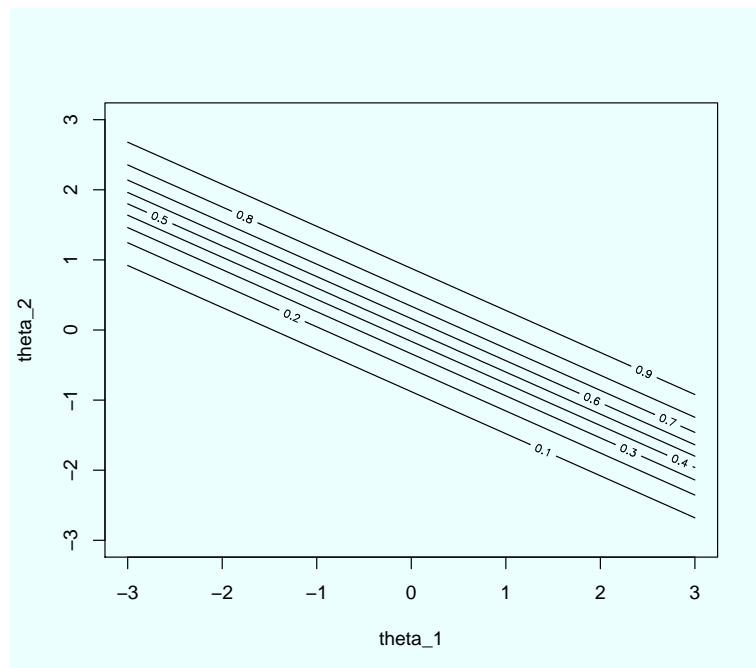


図 2.7: 2 パラメータ・ロジスティック補償型モデルの項目特性曲面の等高線
プロットの例

確認した (2.3) 式の 1 次元の IRT の 2 パラメータ・ロジスティックモデルは、 $d_j = -a_j b_j$ とすれば、

$$p(x_j = 1 | \theta, a_j, b_j) = \frac{\exp(a_j(\theta - b_j))}{1 + \exp(a_j(\theta - b_j))} \quad (2.12)$$

$$= \frac{\exp(a_j\theta - a_j b_j)}{1 + \exp(a_j\theta - a_j b_j)} = \frac{\exp(a_j\theta + d_j)}{1 + \exp(a_j\theta + d_j)} \quad (2.13)$$

と表現できる。これは、1 次元の 2 パラメータ・ロジスティックモデルが多次元 IRT の補償型モデルにおいて次元の数が 1 である場合に相当することを示すものであり、補償型の多次元 IRT のモデルは一次元のロジスティックモデルの自然な拡張となっていることが理解できる。

また、上の関係から、多次元 IRT の困難度に関するパラメータ d_j は、符号やスケールが 1 次元の場合とは異なっていることがわかる。これを調整した多次元項目困難度 $MDIFF$ が補償型の多次元 IRT の場合の項目困難度の指標として提案されている (Reckase, 2009)。

$$MDIFF_j = \frac{-d_j}{\sqrt{\mathbf{a}_j^T \mathbf{a}_j}} \quad (2.14)$$

2.2.2 2 パラメータ・ロジスティック部分補償型モデル

部分補償型の 2 パラメータ・ロジスティックモデル (Two-Parameter Logistic Partial-compensatory Mode, Reckase, 2009) は、次式で表される。

$$P(x_j = 1 | \theta, \mathbf{a}_j, \mathbf{d}_j) = \prod_{l=1}^m \frac{\exp(a_{jl}\theta_l + d_{jl})}{1 + \exp(a_{jl}\theta_l + d_{jl})} \quad (2.15)$$

ただし、上式の m は次元数を表す。補償型のモデルでは、受検者特性値の各次元が和の形で正答確率に影響することを想定していたが、部分補償型モデルでは次元毎に 1 次元のロジスティックモデルと同じ形を想定し、その積の形で正答確率をモデル化する。したがって、項目に関連するパラメータ a と d は次元毎にそれぞれ異なった値を取ることが想定される。

部分補償型モデルは、項目への正答のためにすべての次元の能力特性が高いことを要求するようなテスト、逆にいえば 1 つでも水準が低い能力特性がある場合には正答確率が大きく低下するようなテストの分析を想定したモデルであると捉えることができる。

2.2.3 多次元段階反応モデル

1 次元の場合の段階反応モデルの潜在特性を多次元に拡張したモデルとして多次元段階反応モデル (Muraki & Carlson, 1995) が提案されている。この

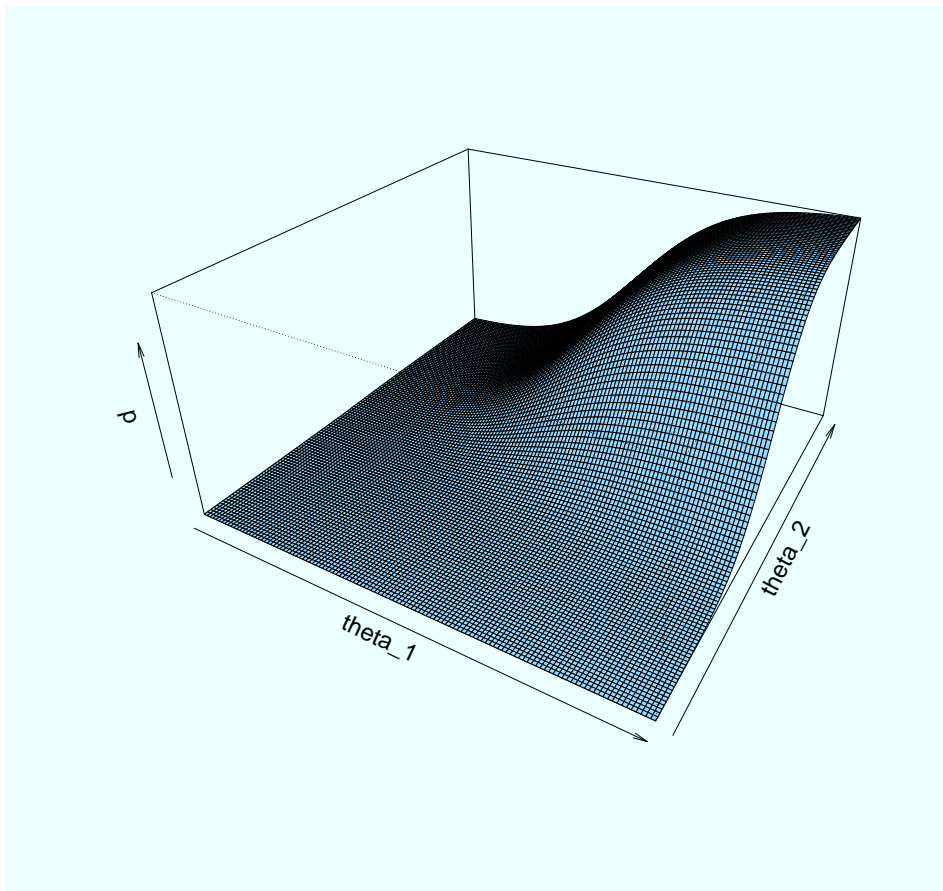


図 2.8: 2 パラメータ・ロジスティック部分補償型モデルの項目特性曲面の例

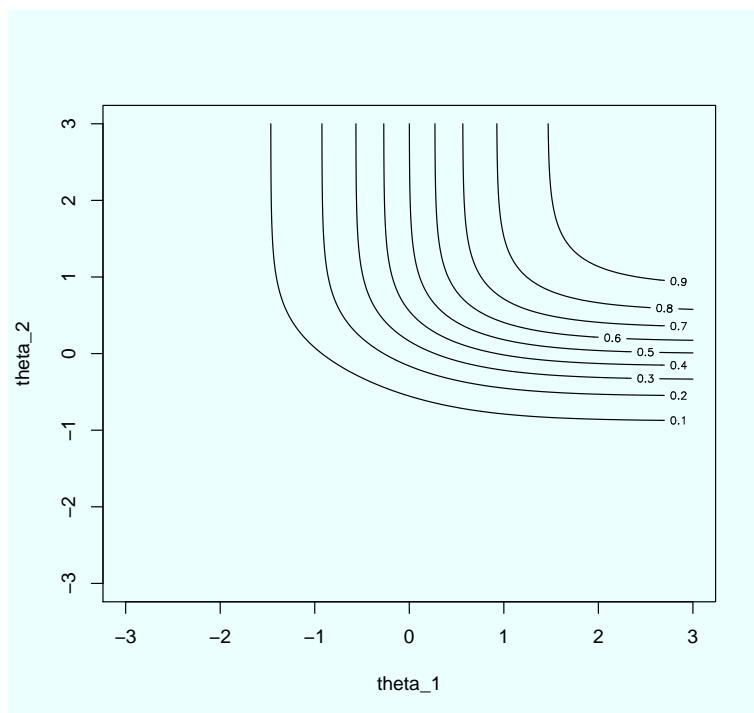


図 2.9: 2 パラメータ・ロジスティック部分補償型モデルの項目特性曲面の等高線プロットの例

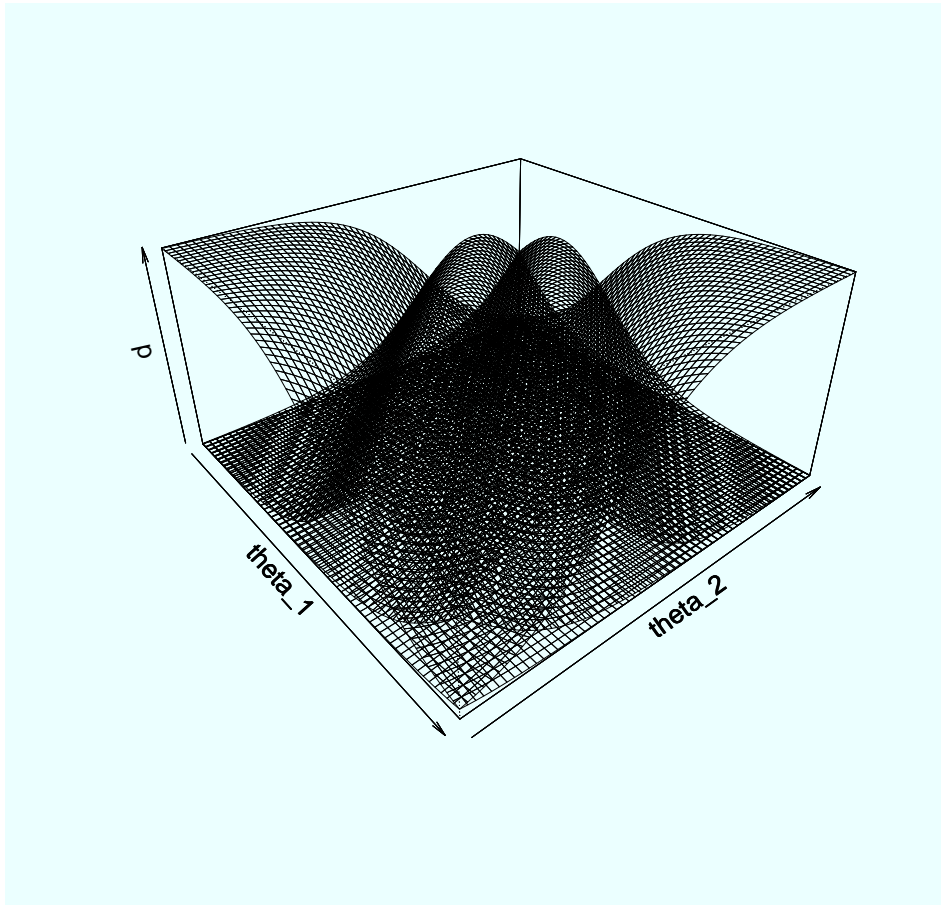


図 2.10: 多次元段階反応モデルの項目特性曲面の例

モデルでは, 1次元の段階反応モデルと同様に, 項目 j においてカテゴリが k となる確率を次式のようにモデル化する.

$$P_{jk}(\boldsymbol{\theta}) = P_{jk}^+(\boldsymbol{\theta}) - P_{j,k+1}^+(\boldsymbol{\theta}) \quad . \quad (2.16)$$

多次元段階反応モデルの項目特性曲面を例として描画したものが次の図 (2.10) である.

上式の $P_{jk}^+(\boldsymbol{\theta})$ は項目 j において k 以上の段階が選ばれる確率を表し, 1次元の段階反応モデルの場合と同様 $k = 0$ の場合と $k = K_j + 1$ の場合について

$$P_{j0}^+(\boldsymbol{\theta}) = 1, \quad P_{j,K_j+1}^+(\boldsymbol{\theta}) = 0 \quad (2.17)$$

とする.

他の $P_{jk}^+(\boldsymbol{\theta})$ には, つぎの正規累積モデルを仮定する.

$$P_{jk}^+ = \Phi(\mathbf{a}_j^T \boldsymbol{\theta} + d_{jk}) \quad . \quad (2.18)$$

ただし、上式の $\Phi(z)$ は標準正規分布の累積密度関数で、標準正規分布の密度関数を ϕ とした場合、

$$\Phi(z) = \int_{-\infty}^z \phi(t) dt \quad (2.19)$$

である。 \mathbf{a} パラメータは、段階 k に依らず各項目 j については共通した値のベクトルとなっていることに注意する。

以上、多次元 IRT の 3 つのモデルについて確認したが、これ以外の多次元 IRT のモデルとしては、1 つでも基準を超える特性次元があれば正答確率が高くなる状況を表現した OR モデル (藤森, 1985; 藤森, 1988)、補償型モデルと部分補償型モデルを統合したモデル (岡田・前川, 2014) などが提案されている。

これらのモデルのうち最もよく取り上げられ研究されているのは正誤 2 値データに対応した 2 パラメータ・ロジスティック補償型モデルであり、単に多次元 IRT という場合は、通常補償型のロジスティックモデルのことを指す (室橋, 2013)。本研究でも、以降はロジスティック補償型のモデルについて検討する。

2.2.4 多次元 IRT と因子分析モデルの関係

多次元 IRT は、観測変数である項目に対する反応データを、潜在変数である受検者特性値と項目特性値によって表現する統計モデルである。ところで、複数の観測変数と多次元の潜在特性の関連性を捉える代表的な統計モデルとして因子分析モデルがある。Takane & Leeuw(1987) は、補償型の多次元 IRT のモデルが観測変数が離散的な値を取る場合の因子分析モデルであると見なせることを明らかにしている⁵。ここでは、この関係をパス図を利用して簡単に確認する。

因子分析のモデルは、次式で表される。

$$x_j = \mathbf{a}_j^T \boldsymbol{\theta} + d_j + e_j \quad . \quad (2.20)$$

上式で、 x_j は変数 j に対応する実数値である。 $\boldsymbol{\theta}$ は潜在変数ベクトルを表すもので、因子スコアと呼ばれ、平均 0 の多変量正規分布に従うことが仮定されることが多い。また、 \mathbf{a}_j は上式を $\boldsymbol{\theta}$ を独立変数とした回帰モデルとみなした場合の回帰係数に相当するベクトルで、因子負荷量と呼ばれる。 d_j は切片項である。また、 e_j は誤差項であり、通常 平均 0 の独立な正規分布に従うことが

⁵他に、Reise *et al.*, 2011 など参照。

仮定される。因子分析の特徴は、複数の観測変数の特徴を、比較的少数の潜在変数 θ を用いて表すことにある。上式は、誤差項の存在を除けば、多次元 IRT の補償型ロジスティックモデルの指数関数の中の部分とほとんど同じ形をしていることがわかる。

図 (2.11) は、因子分析のパス図の例である。また、図 (2.12) は補償型の 2 パラメータ・ロジスティックモデルのパス図の例である⁶。因子分析の場合には因子スコアが、多次元 IRT の場合には受検者特性値が潜在変数であるが、それぞれの図中の楕円はこれを表現している。楕円間の双方向の矢印は、潜在変数間の相関を表したものである。また、四角形は観測変数を表しており、因子分析の場合には実数値、多次元 IRT の場合には項目への離散的な反応がこれにあたる。多次元 IRT の図 (2.12) において観測変数を表す四角形に横線が入っているのは、補償型の多次元 IRT モデルにおける観測変数が、ある種の閾値を伴う変数の離散化の結果として解釈できることを表現したものである。観測変数から潜在変数への片方向の矢印は、因子分析の場合は因子負荷量、多次元 IRT の場合には識別力に相当する a が 0 以外の値であることを表している。矢印が存在しない場合には当該の因子負荷量あるいは識別力が 0 であることを意味する。

2 つの図を見比べると、因子分析では明示的な誤差項が存在するが、多次元 IRT ではこれが存在しない点が大きな違いであることがわかる。また、多次元 IRT では観測変数を潜在変数が離散的に変換されたものとして捉えるが、因子分析では観測変数は潜在変数と直接に線形の関係をもつものとして理解される。

これらの点を除けば、補償型の多次元 IRT は因子分析モデルと類似する部分が多い。前に多次元 IRT モデルが 1 次元の IRT モデルの自然な拡張となっていることを確認したが、別の視点として多次元 IRT モデルをデータが離散値の場合の因子分析モデルとして捉えることが可能である。ただし、因子分析では主に観測変数群を少数の潜在変数で説明することが分析の目的であり、因子負荷量の解釈が中心的な興味となる。したがって因子スコアへの注目の度合いは通常それほど高くはない。一方、多次元 IRT では推定された受検者特性値がテストのスコアとしての意味をもつため、分析結果において受検者特性値すなわち因子スコアが重視される点が、両モデルの応用上の大きな差異となっている点に注意する必要がある。

⁶Rupp *et al.* (2010) を参考にした。

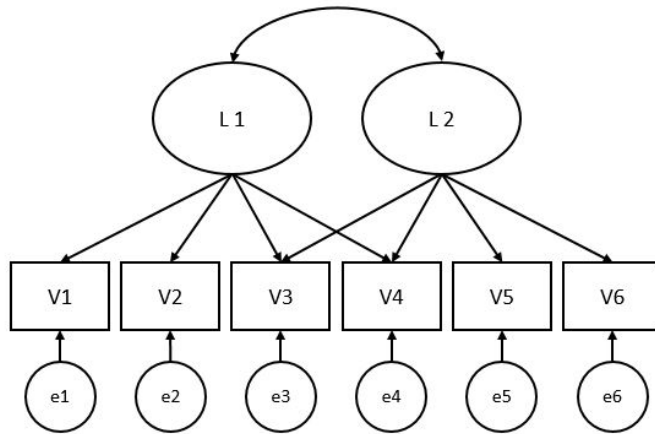


図 2.11: 因子分析モデルのパス図の例

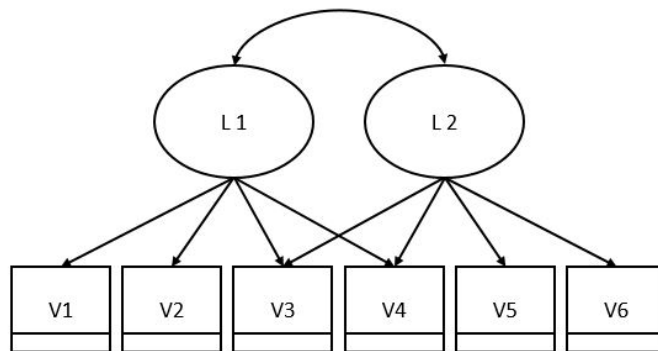


図 2.12: 多次元 IRT 補償型 2 パラメータ・ロジスティックモデルのパス図の例

第3章 多次元IRTの等化の先行研究

IRT の重要な研究領域の一つに複数のテストのスコアを同一の基準の上で比較可能とする等化法に関する研究がある。IRT は、問題項目の特性と受検者の特性を分離して把握することで、問題項目に依存しない形で受検者の特徴を把握することを目指すものであるが、等化の手続きはこの趣旨において、IRT の研究および実用場面での活用にとっての重要な位置づけを占める。

各種の等化法については、たとえば本邦においても積極的な研究が進められている (e.g. Arai & Mayekawa, 2011, 荒井・菊池・前川, 2013, 泉・山野井・山田・金森・対馬, 2012, 光永・前川, 2013, 光永・御園・水町, 2013)。一方で、国内、国外を含めて等化法に関する既存の研究のほとんどは 1次元の IRT に関するものである。多次元 IRT の等化法は、比較的新しい研究分野であり (Simon, 2008), まだ十分な研究が行われていないとの指摘がある (Min, 2007)。

さて、多次元 IRT においても、1次元の IRT の場合と同様に、等化が可能な条件として複数のテストフォームに共通して含まれる項目群である共通項目が存在する場合と、複数のテストフォームに共通した受検者である共通受検者が存在する場合が考えられる (Angoff, 1982)¹。これらの場合には、テストフォームに共通する項目や受検者をアンカー (錨) として利用することで、各テストフォームの受検者が同じスケール上に布置するように変換を行う。この操作が等化と呼ばれる。

ところで等化は通常、項目パラメータや受検者特性値が、テストフォーム毎にあらかじめ計算されている状況 (separate caribration) において、共通項目や共通受検者を利用して、テストフォーム毎に推定されたパラメータを同じスケールで表現することを目指す手続きを指す。これとは別に、テストフォーム毎に項目パラメータや受検者特性値を計算するのではなく、構造的にデータが存在しない部分を欠測値として、最初から全てのテストフォームの項目パラメータや受検者特性値を一度に推定することを目指す方法 (concurrent caribration) が存在するが、本研究では、専ら前者の場合を想定する。また、多次元 IRT の等化法は、補償型のロジスティックモデルを中心に研究されているため、本研究でも、このモデルに関する等化法について検討する。

以降では、まず多次元 IRT の等化の前提となるパラメータの不定性について確認し、その後、先行研究で示された主な等化法について確認する。

3.1 多次元 IRT のパラメータの不定性

いま、(2.11) 式の指数関数の中の部分を取り出して $z_j = \mathbf{a}_j^T \boldsymbol{\theta} + d_j$ とし、これをテストを構成する項目および受検者についてまとめた次の行列について考える。

$$Z = \Theta A^T + \mathbf{1}_n \mathbf{d}^T \quad . \quad (3.1)$$

¹これ以外に Reckase(2009) では問題項目の異なる 2 つのテストの受検者群が同じ母集団からランダムにサンプリングされた場合が紹介されているが、多次元 IRT ではこの設定での等化は不安定になるため避けるべきであるとの指摘がある (Davey *et al.*, 1996)。

上式右辺の Θ は θ^T を受検者数分縦に並べた $n \times m$ の行列, A は \mathbf{a}_j^T を項目数分縦に並べた $p \times m$ の行列, $\mathbf{1}_n$ は 1 を n 個並べたベクトル, \mathbf{d} は d_j を項目数分縦に並べた要素数が p のベクトルである. なお, m は次元の数, p は項目数, n は受検者数をそれぞれ表す. さて, (3.1) 式の Θ , A , \mathbf{d} について, 任意の正則な $m \times m$ の行列 W と要素数が m のベクトル \mathbf{u} を利用して, それぞれ次のように変換することを考える.

$$\tilde{\Theta} = \Theta W + \mathbf{1}_n \mathbf{u}^T, \quad (3.2)$$

$$\tilde{A}^T = W^{-1} A^T, \quad (3.3)$$

$$\tilde{\mathbf{d}} = \mathbf{d} - A(W^{-1})^T \mathbf{u} . \quad (3.4)$$

このとき, (3.1) 式の Θ , A , \mathbf{d} をそれぞれ $\tilde{\Theta}$, \tilde{A} , $\tilde{\mathbf{d}}$ と入れ替えても

$$\begin{aligned} \tilde{Z} &= \tilde{\Theta} \tilde{A}^T + \mathbf{1}_n \tilde{\mathbf{d}}^T \\ &= (\Theta W + \mathbf{1}_n \mathbf{u}^T) W^{-1} A^T + \mathbf{1}_n [\mathbf{d} - A(W^{-1})^T \mathbf{u}]^T \\ &= \Theta A^T + \mathbf{1}_n \mathbf{d}^T \end{aligned}$$

となって (3.1) 式に一致する. すなわち, 2 パラメータ・ロジスティック補償型モデルの多次元 IRT では, 受検者特性値 Θ と項目パラメータ A , \mathbf{d} の間に一種の不定性が存在し, 双方に上記のような変換を行なっても正答確率には影響を与えないことがわかる.

3.2 多次元 IRT の等化

上記の不定性を利用することで, 多次元 IRT の等化を実現することができる. まず, 共通受検者による等化についてみると, 同じ特性次元を測定する 2 つのテストフォームについて, 両方に共通する受検者が存在し, それらの特性値がテストフォームを通じて変わらないことを仮定できる場合に, 前述の不定性を利用してテストフォームを等化することができる. 等化先のテストにおける受検者特性値 θ^T を n 人分縦に並べた行列を Θ , 等化対象となるテストにおける受検者の特性値 θ_g を同様にまとめた行列を Θ_g と記すことにする. なお Θ_g の各行はそれぞれ Θ と対応しているものとし, Θ と Θ_g の関係に次式を仮定する.

$$\Theta = \Theta_g W_g + \mathbf{1}_n \mathbf{u}_g^T . \quad (3.5)$$

上式の \mathbf{u}_g は Θ_g の位置の変換に対応するベクトル, W_g は回転 (rotation), 伸縮 (dilation), 次元の入れ替え (reflection) の変換を表す等化係数行列である. (3.5) 式の関係は通常厳密には満たされないが, $\hat{\Theta} = \Theta_g W_g + \mathbf{1}_n \mathbf{u}_g^T$ として, Θ と $\hat{\Theta}$ ができるだけ一致するように W_g などのパラメータを調整することで等化を実現する.

共通項目を利用した等化の場合には、等化先のテストにおける共通項目の \mathbf{a} パラメータを転置したベクトルを縦にならべた行列を A , d パラメータを縦にならべたベクトルを \mathbf{d} とし、等化対象となるテストフォームにおける同じ項目群の行列を A_g , \mathbf{d}_g とする。 A_g の各行は A と、 \mathbf{d}_g の各要素は \mathbf{d} とそれぞれ対応しているものとし、次式の間係を仮定する。

$$A^T = W_g^{-1} A_g^T \quad , \quad (3.6)$$

$$\mathbf{d} = \mathbf{d}_g - A_g(W_g^{-1})^T \mathbf{u}_g \quad . \quad (3.7)$$

上式の間係はやはり通常厳密には満たされないが、 $\hat{A}^T = W_g^{-1} A_g^T$, $\hat{\mathbf{d}} = \mathbf{d}_g - A_g(W_g^{-1})^T \mathbf{u}_g$ として、 A と \hat{A} , \mathbf{d} と $\hat{\mathbf{d}}$ をできるだけ近づけるように W_g などのパラメータを調整することで等化を実現する。

3.3 多次元 IRT の等化法 (先行研究)

ここで、多次元 IRT について先行研究で示された主な等化法について確認する。なお、各等化法の特徴について表 (3.1) にまとめる。

3.3.1 TCF 法

TCF 法 (Oshima *et al.*, 2000) は、1次元の IRT の Stocking & Lord の方法 (1983) を多次元 IRT に拡張した等化法である。TCF 法では、次式で表される基準 Q_{SL} を最小化する等化係数を求める (本橋, 2013)。

$$Q_{SL} = \int \dots \int_{-\infty}^{\infty} \left(\sum_{j=1}^p P_j(\boldsymbol{\theta})^{(1)} - \sum_{j=1}^p P_j(\boldsymbol{\theta})^{(2^*)} \right)^2 \phi(\boldsymbol{\theta}^{(1)}) d\boldsymbol{\theta}^{(1)} \quad (3.8)$$

$$+ \int \dots \int_{-\infty}^{\infty} \left(\sum_{j=1}^p P_j(\boldsymbol{\theta})^{(1^*)} - \sum_{j=1}^p P_j(\boldsymbol{\theta})^{(2)} \right)^2 \phi(\boldsymbol{\theta}^{(2)}) d\boldsymbol{\theta}^{(2)} \quad (3.9)$$

上式の p は共通項目の数、 $\phi(\boldsymbol{\theta})$ は受検者特性値の密度分布を表す。また、上付きの数値 (1), (2) は、それぞれ等化対象の2つのテストを表し、記号 * は等化係数による変換を行った後の状態で値を計算することを示すものである。TCF 法は、共通項目の項目特性曲面をテストフォーム全体として、できるだけ一致させることを目指すものであり、共通項目による等化法に分類される。

3.3.2 ICF 法

ICF 法 (Oshima *et al.*, 2000) は、1次元の IRT の Haebara 法 (1980) を多次元 IRT に拡張した等化法であり、次の基準 Q_H を最小化する等化係数行列

を求める.

$$Q_H = \sum_{j=1}^p \left[\int \dots \int_{-\infty}^{\infty} \left(P_j(\boldsymbol{\theta})^{(1)} - P_j(\boldsymbol{\theta})^{(2^*)} \right)^2 \phi(\boldsymbol{\theta}^{(1)}) d\boldsymbol{\theta}^{(1)} \right] \quad (3.10)$$

$$+ \sum_{j=1}^p \left[\int \dots \int_{-\infty}^{\infty} \left(P_j(\boldsymbol{\theta})^{(1^*)} - P_j(\boldsymbol{\theta})^{(2)} \right)^2 \phi(\boldsymbol{\theta}^{(2)}) d\boldsymbol{\theta}^{(2)} \right] \quad (3.11)$$

p , $\phi(\boldsymbol{\theta})$ および, 上付きの数値 (1), (2) 等の意味するところは, TCF 法と同様である. ICF 法は, 2つのテストフォームに含まれる個々の共通項目の項目特性曲面を, 全体としてできるだけ一致させることを目指すものであり, やはり共通項目による等化法に分類される.

3.3.3 Li & Lissitz の方法

Li & Lissitz の方法 (Li & Lissitz, 2000) は, 因子分析におけるプロクラステスト回転を応用した方法である. 2つのテストフォームに含まれる共通項目について, 一方のテストフォームを基準とし, もう一方のテストフォームの共通項目の項目パラメータをこれに近づけることを考える. 基準となる等化先のテストフォームの共通項目の \mathbf{a} パラメータ行列を A , d パラメータ・ベクトルを \mathbf{d} , 等化対象となるテストフォームの共通項目の \mathbf{a} パラメータ行列を A_g , d パラメータ・ベクトルを \mathbf{d}_g とする. また, A と A_g をそれぞれ中心化した行列を A_c , A_{gc} とする.

Li & Lissitz の方法では, 回転行列 W_g について

$$W_g^{-1} = kT^T \quad (3.12)$$

なる構造を考える. k は \mathbf{a} パラメータのスケールを調整するスカラー値, T は $T^T T = T T^T = I$ なる直交回転行列である.

また,

$$A_c^T = (W_g^{-1})^T A_{gc}^T \quad (3.13)$$

$$\mathbf{d} = \mathbf{d}_g - A_{gc}(W_g^{-1})^T \mathbf{u}_g \quad (3.14)$$

であると考え, この関係をできるだけ満たすような等化係数行列 W_g , \mathbf{u}_g を求める.

なお, これらの等化係数について, TCF 法や ICF 法のように項目特性曲面を利用して一度に求めることも可能ではあるが, 精度の面からそれぞれを順に計算する方法が推奨されている (Li & Lissitz, 2000; 本橋, 2013).

手順として, まず

$$\text{tr}[(A_c - A_{gc}T)^T (A_c - A_{gc}T)] \quad (3.15)$$

を最小化する T を直交プロクラステス回転 (Shönemann,1966) によって求める。つぎに, k を

$$k = \text{tr}[(A_{gc}T)^T A_c] / \text{tr}[A_{gc}^T A_c] \quad (3.16)$$

によって求め, さらに

$$\|\mathbf{d} - (\mathbf{d}_g - A_{gc}(W_g^{-1})^T \mathbf{u}_g)\|^2 \quad (3.17)$$

を最小化する \mathbf{u}_g を求める。等化のための回転操作については直交性を仮定し, 次元に共通したスカラー k でスケールを調整する点が Li & Lissitz の方法の特徴である。

3.3.4 Min の方法

Min(2003, 2007) は, つぎの方法で共通項目を利用して等化を行うことを提案している。まず, T を直交回転行列, K を $m \times m$ の対角行列とし, 回転行列 W_g について

$$W_g^{-1} = (TK)^T \quad (3.18)$$

とする²。Li & Lissitz の方法では, \mathbf{a} パラメータのスケールを調整する k はスカラーであったが, Min の方法は, これを対角行列とするものである。

等化のためには,

$$\text{tr}[(A - A_g T)^T (A - A_g T)]$$

を最小化する直交回転行列 T を求め, 次に K を

$$K = \text{diag}(A^T A_g T) [\text{diag}(T^T A_g^T A_g T)]^{-1}$$

として, (3.18) 式の W_g を求める (Simon, 2008)。

多次元 IRT における \mathbf{a} パラメータの等化は, 軸の回転 (rotation), 伸縮 (dilation), 次元の入れ替え (reflection) が複合したものとして捕らえられるが, 上の T は回転と次元の入れ替えに, K は次元毎の伸縮に対応したものである。Min の方法は, 軸の回転については直交回転とし, 軸の伸縮を各次元毎に取り扱う点に特徴がある。

3.3.5 NOP 法

Yon(2006) による NOP 法 (Non-Orthogonal Procrustes Method) は, Min の方法と類似した方法であるが, 回転行列に直交性を仮定しない点, 伸縮に対応した K 行列を利用しない点が異なる。

²各等化法における W_g, T は異なった値を取る行列であるが, 比較のために同じ表記とした。

NOP 法の回転行列 W_g は,

$$W_g^{-1} = (A_g^T A_g)^{-1} A_g^T A_g \quad (3.19)$$

で与えられる (Simon, 2008). NOP 法は, A を従属変数群, A_g を独立変数群とした, 制約のない多変量回帰であると解釈できる.

Min の方法などの直交回転に基づく方法では, 等化後のパラメータ行列 A の値が, 等化先のそれに十分に近接できない場合がある. これは, 直交性が一種の制約として働くためであるが, 斜交回転に基づく方法の場合, この制約が緩和されるため, 等化による追従性が高くなることが期待できる. NOP 法は, このような効果を狙った方法であり, 特に次元数が高い場合に有効であるとされている (Yon, 2006).

3.3.6 Hirsch の方法

Hirsch(1989) は, 次の方法で共通受検者を利用して W_g を求めることを提案している. まず Θ と Θ_g の基底を揃えるために Θ_g を

$$\Theta_g^* = \Theta_g P \Lambda^{-1/2} \Delta^{1/2} Q^T$$

と変換する. ただし, Λ, P はそれぞれ θ_g の分散共分散行列 S_g をスペクトル分解した場合の固有値を対角に並べた行列と, 対応する固有ベクトルを順に並べた行列, Δ, Q は θ の分散共分散行列 S をスペクトル分解した場合の固有値を対角に並べた行列と固有ベクトルを順に並べた行列であり, $S_g = P \Lambda P^T$, $S = Q \Delta Q^T$, $P P^T = P^T P = Q^T Q = Q Q^T = I$ である. つぎに, 変換後の Θ_g^* に Θ をターゲットとした直交プロク拉斯テス回転を適用することによって回転行列 T を求める. T は軸の回転と次元の入れ替えを担う直交行列である. T が得られれば W_g は,

$$W_g = P \Lambda^{-1/2} \Delta^{1/2} Q^T T$$

によって求めることができる.

ここで, θ_g の等化後の分散共分散行列について調べておく. 等化後の θ_g の分散共分散行列は

$$\begin{aligned} & W^T S_g W \\ &= (P \Lambda^{-1/2} \Delta^{1/2} Q^T T)^T S_g (P \Lambda^{-1/2} \Delta^{1/2} Q^T T) \\ &= T^T S T \end{aligned}$$

となる. もし $S = I$ であれば上式は $T^T S T = I$ となるが, θ の次元間には特別な場合を除いて相関があるため, 一般には $S = I$ ではない. したがって Hirsch の方法によって等化を行った場合の変換後の能力特性値の分散共分散

行列は、通常等化先の能力特性値のものとは一般には一致しない。

上記以外の多次元 IRT の等化法としては、たとえば 1 次元の IRT の場合の mean/mean 法や mean/sigma 法を拡張した方法 (Davey *et al.*, 1996; Oshima *et al.*, 2000), 2 つのテストフォームの a パラメータと d パラメータについてそれぞれの差の二乗和を加えた値を最小化する直接法 (Divgi, 1985; Oshima *et al.*, 2000), 直交プロクラステス回転による方法 (Tompson *et al.*, 1997) などの方法が提案されているが、いずれも共通項目による等化法である。

以上のように、多次元 IRT の等化に関して、共通項目を利用した場合については、いくつかの方法が提案されている。一方、共通受検者による等化法に関する研究は Hirsch (1989) によるものの他にはほとんど見当たらないのが現状である。この要因として、共通受検者による等化では、受検者の能力特性値が、2 つのテストの受検時点で変化しないことを仮定する必要があるが、共通項目による等化に比較して利用上の制約が大きいため、どちらかといえば共通項目による等化のほうが設計が容易である点が考えられる。しかしながら実務上、共通受検者デザインによる等化のほうが適切な場面が少ないわけではない。また、後にみるように、多次元 IRT においては特に、それまで 1 次元の IRT を利用して運用してきた複数のテストを多次元 IRT の枠組みで統合するような利用法が考えられるが、この場合には、共通受検者による等化が有効な可能性が高い。

ところで、Hirsch の方法は共通受検者による等化法として最良のものと捉えられるかという点については、そうではないように思われる。Hirsch の方法の問題として、つぎの 2 点を指摘できる。まず、Hirsch の方法による等化の操作は、基本的には直交回転に基づくものであるが、回転における直交性は一種の制約として働くため、この方法では等化後の 2 つのテスト・フォームの共通受検者の特性値の近接性が十分に高くならない可能性がある。

Hirsch の方法のもう一つの課題として、この方法では等化後の 2 つのテスト・フォームの共通受検者の次元間の相関係数が、必ずしも一致しないことが挙げられる。テストの次元間の相関係数は、テストの特性を表す重要な量であり、等化先のテストと等化後のテストでこれが異なることは、スコアの解釈という面において支障を生じる可能性がある。また、次元毎のスコアの荷重和による合成スコアを各受検者の総合スコアとするような場合に、次元間の相関がテスト・フォーム毎に異なることは、このスコアの実質的な意味がテスト・フォーム毎に異なることにつながる³。したがって、等化後の受検者特性値の次元間相関は、等化先のものとは一致していることが望ましい。

これらの事柄を鑑みると、多次元 IRT の共通受検者による新たな等化法を

³付録 1 に次元間相関が合成スコアに影響を与える例を示した

提案することの意義は小さくないと考えられる。そこで、本研究では上記の課題を踏まえた上で、共通受検者を利用した複数の等化法を開発することを目指す。

表 3.1: 先行研究で示された主な等化法

共通項目/共通受検者	等化法	特徴
共通項目	TCF 法 (Oshima <i>et al.</i> , 2000)	一次元の IRT の場合の Stocking & Lord の方法を多次元 IRT に拡張した方法
	ICF 法 (Oshima <i>et al.</i> , 2000)	一次元の IRT の場合の Haebara 法を多次元 IRT に拡張した方法
	Li & Lissitz の方法 (Li & Lissitz, 2000)	直交プロクラステス回転に基づく多次元 IRT に固有の方法
	Min の方法 (Min, 2003)	Li & Lissitz のスケールの調整を次元毎に行う形に修正した方法
	NOP 法 (Yon, 2006)	斜交プロクラステス回転による方法
共通受検者	Hirsch の方法 (Hirsch, 1989)	受検者特性値に注目した直交プロクラステス回転に基づく方法

第4章 共通受検者を利用した多次元IRTの等化法

4.1 問題

前章で確認したとおり, 多次元 IRT の共通受検者による等化法は Hirsch によるものを除いてほとんど提案されていない. また必ずしも Hirsch の方法が最良な共通受検者による等化法であるというわけではないと考えられる. この章では, 多次元 IRT における共通受検者を利用した場合の新たな等化法として, 因子分析における回転法を応用した 2 つの方法を示す.

なお, 通常因子分析では, 因子負荷量の解釈が中心的な興味であり, 回転操作は専ら因子負荷量に注目して構成されている. 一方, ここで検討する多次元 IRT における共通受検者による等化は, 因子分析の文脈では因子スコアに注目した回転に相当する. 因子負荷量に注目した場合の回転解と因子スコアに注目した回転解では回転の基準が異なるため, 多次元 IRT の共通受検者による等化に因子分析の回転を応用する場合には, 因子スコアによる回転に対応する解を導く必要がある. 本章では多次元 IRT の新たな等化法として 2 つの基準を提案し, それぞれに対応した解を示す. また, これらの方法に直交プロクラステス回転法, Hirsch の方法を加えた 4 つの方法について計算機によるシミュレーション実験を行って性質を検証する¹.

4.2 提案手法

さて, Hirsch の方法は 2 つのテストフォームの共通受検者の特性値について, 基底を一致させてから直交回転を行うことで等化を実現するものであった. この手法の問題として, 前章でつぎの 2 点を指摘しておいた. まず, Hirsch の方法による等化の操作は, 基本的には直交プロクラステス回転に基づいて行われるが, 回転における直交性は一種の制約として働くため, この方法では等化後の 2 つのテストフォームの共通受検者の特性値の近接性が十分に高くない可能性がある点である². したがって, 望ましい等化を実現する一つの方向性として, 共通項目による等化の場合の NOP 法 (Yon, 2006) と同様, より制約の少ない斜交回転に基づく方法を利用することが考えられる.

また, Hirsch の方法のもう一つの課題として, 等化後の 2 つのテストフォームの共通受検者の次元間の相関係数が, 必ずしも一致しないという問題がある. テストの次元間の相関係数は, テストの特性を表す重要な量であるから, 等化後の受検者特性値の次元間相関は, 等化先のものと同じであることが望ましい. この方向の Hirsch の方法の改良として, 相関係数に一致制約を課

¹共通受検者による直交プロクラステス回転法は, 多次元 IRT の等化法として直接には提案されたものではないが, Hirsch の方法による等化のプロセスの中で利用される計算手法の一部を取り出したものであり, 因子スコアに注目した直交プロクラステス回転に相当する. 具体的な計算法については付録に示す.

²ただし, 基底を一致させる操作を含めて考えると, Hirsch の方法は斜交回転法の一つとして捉えられる.

した等化法を考えることができる。このアプローチは、後で示すように因子分析の場合の重み付きプロクラステス回転法に帰着する。

以下では、上記のそれぞれの方向性に対応した方法として、受検者特性値に注目した斜交プロクラステス回転による方法、同じく受検者特性値に注目した重み付きプロクラステス回転法の2つの方法について示す。

4.2.1 受検者特性値の一致度を高める方法

受検者特性値の一致度をより高める方法として、因子分析の場合の斜交プロクラステス回転を応用することが考えられる。いま、等化先のテストにおける受検者特性値 $\boldsymbol{\theta}^T$ を n 人分縦に並べた行列を Θ 、等化対象となるテストにおける受検者の特性値 $\boldsymbol{\theta}_g$ を同様にまとめた行列を Θ_g と記すことにする。なお Θ_g の各行はそれぞれ θ と対応しているものとし、 Θ と Θ_g の関係に次式を仮定する。

$$\Theta = \Theta_g W_g + \mathbf{1}_n \mathbf{u}_g^T \quad (4.1)$$

上式の \mathbf{u}_g は Θ_g の位置の変換に対応するベクトル、 W_g は回転 (rotation)、伸縮 (dilation)、次元の入れ替え (reflection) の変換を表す等化係数行列である。上式の関係は通常厳密には満たされないが、 $\hat{\Theta} = \Theta_g W_g + \mathbf{1}_n \mathbf{u}_g^T$ として、 Θ と $\hat{\Theta}$ ができるだけ一致するように W_g などのパラメータを調整することで等化を実現する。なお、以降簡単のために、 Θ と Θ_g は共に各列の平均が0、各次元の分散が1に標準化されているものとする。このとき $\mathbf{u}_g = \mathbf{0}$ である。したがって、この場合の等化の手続きは、何らかの基準において Θ と $\hat{\Theta}$ の距離を最小化する等化係数行列 W_g を求めることに対応する。

さて、共通受検者による等化を、等化係数行列 W_g による変換によって等化先の受検者特性値の行列 Θ に等化対象のテストの受検者特性値行列 Θ_g をできるだけ近接させるものと捉えれば、この問題は次式のように定式化できる。

$$\min f = \frac{1}{2} \text{tr} [(\Theta - \Theta_g W_g)^T (\Theta - \Theta_g W_g)] \quad (4.2)$$

$$\text{s.t. } \text{diag}(W_g^T S_g W_g) = I \quad (4.3)$$

S は等化先の受検者特性値の次元間の分散共分散行列を、 S_g は等化先の受検者特性値の次元間の分散共分散行列を表す。また、 W_g は回転行列である。制約条件は、等化後の各次元の分散が1であるとするもので、等化後の受検者特性値の分散が等化先と一致することを要請するものである。

たとえば、共通項目による等化法のひとつである NOP 法では、上記のような受検者特性値の分散に関する制約については考慮されていないが、受検者特性値による等化においては、この種の制約について考慮する必要がある³。

³この制約は、因子分析の場合には一般的なものである。また Hirsch の方法においても、厳密ではないが同様の制約を考慮した構成となっている。

さて、等化係数行列 W_g を斜交回転行列であると考え、(4.3) 式の制約を考慮せずに (4.2) 式を最小化する W_g を求めると

$$W_g = \frac{1}{n} S_g^{-1} \Theta_g^T \Theta \quad (4.4)$$

となる。このとき、 θ_g の変換後の分散共分散行列は、

$$\begin{aligned} W_g^T S_g W_g &= \left(\frac{1}{n} S_g^{-1} \Theta_g^T \Theta \right)^T S_g \left(\frac{1}{n} S_g^{-1} \Theta_g^T \Theta \right) \\ &= \frac{1}{n^2} \Theta^T \Theta_g S_g^{-1} \Theta_g^T \Theta \end{aligned}$$

となるが、この対角成分は必ずしも 1 にはならず (4.3) 式の制約は満たされない。したがって変換後の各次元の分散が 1 となるよう次の調整を行う。

$$W_g^* = W_g [\text{diag}(W_g^T S_g W_g)]^{-1/2} .$$

因子分析における因子負荷量に注目した回転の場合、調整後の回転行列 W_g^* は一種の近似解であり、制約条件下で目的関数を厳密に最小化するものとはならない (市川,2010) が、本研究で取り扱う受検者特性値すなわち因子スコアに注目した回転の場合には、この方法で厳密な解が得られる。この点に関する証明を含めた、受検者特性値に注目した斜交プロクラステス回転法の詳細な解の導出については付録に記す。

4.2.2 相関係数を一致させる方法

Hirsch の方法では、 Θ と Θ_g の基底を一致させてから回転を行なっているが、別の方法として基底の変換と直交回転を同時に実現することが考えられる⁴。この場合、等化は次の基準を最小化する直交回転行列 T を求めることに対応する。このように工夫することで、後述するように変換後の θ_g の分散共分散行列が θ の分散共分散行列に必ず一致する。

$$\begin{aligned} f &= \frac{1}{2} \text{tr} [(\Theta - \Theta_g W_g)^T (\Theta - \Theta_g W_g)] \quad (4.5) \\ &= \frac{1}{2} \text{tr} [(\Theta - \Theta_g P \Lambda^{-1/2} T \Delta^{1/2} Q^T)^T (\Theta - \Theta_g P \Lambda^{-1/2} T \Delta^{1/2} Q^T)] . \end{aligned}$$

ただし、 Λ , P はそれぞれ θ_g の分散共分散行列 S_g をスペクトル分解した場合の固有値を対角に並べた行列と、対応する固有ベクトルを順に並べた行列、 Δ , Q は θ の分散共分散行列 S をスペクトル分解した場合の固有値を対角に並べた行列と固有ベクトルを順に並べた行列であり、 $S_g = P \Lambda P^T$, $S = Q \Delta Q^T$, $P P^T = P^T P = Q^T Q = Q Q^T = I$ である。

⁴柳井・繁耕・前川・市川 (1990) の p.22 および p.97 を参考にした。なお、この方法では Hirsch の方法と同様、途中の展開において直交回転を利用するが、全体としては斜交回転となる。

(4.5) 式は

$$f = \frac{1}{2} \text{tr} \left[\left(\Theta Q \Delta^{-1/2} - \Theta_g P \Lambda^{-1/2} T \right) (\Delta^{1/2} Q^T) \right. \quad (4.6)$$

$$\left. (Q \Delta^{1/2}) \left(\Theta Q \Delta^{-1/2} - \Theta_g P \Lambda^{-1/2} T \right)^T \right] \quad (4.7)$$

$$= \frac{1}{2} \text{tr} \left[(\Theta_L - \Theta_{gL} T) \Delta (\Theta_L - \Theta_{gL} T)^T \right] \quad (4.8)$$

と変形することができる。ただし $\Theta_L = \Theta Q \Delta^{-1/2}$, $\Theta_{gL} = \Theta_g P \Lambda^{-1/2}$ とした。ここで Θ_L , Θ_{gL} は, それぞれ Θ と Θ_g を直交化した行列であると見なせるから, 上式の基準を最小化する T は, Θ と Θ_g をそれぞれ直交化した後に, Δ を重みとして考慮した重み付きプロク拉斯テス回転 (Koschat & Swayne, 1991) となる。 T が得られれば W_g は,

$$W_g = P \Lambda^{-1/2} T \Delta^{1/2} Q^T$$

によって求めることができる。この場合の等化後の θ_g の分散共分散行列は

$$\begin{aligned} W_g^T S_g W_g &= W_g^T P \Lambda P^T V_g \\ &= (P \Lambda^{-1/2} T \Delta^{1/2} Q^T)^T (P \Lambda P^T) (P \Lambda^{-1/2} T \Delta^{1/2} Q^T) \\ &= Q \Delta Q^T = S \end{aligned} \quad (4.9)$$

となって等化先の θ のものと常に一致する。

さて (4.5) 式を展開すると,

$$f = \frac{1}{2} \left[\text{tr}(\Theta_L^T \Delta \Theta_L) - 2 \text{tr}(\Delta \Theta_L^T \Theta_{gL} T) + \text{tr}(\Theta_{gL} T \Delta T^T \Theta_{gL}^T) \right] \quad (4.10)$$

となるが, 上式の右辺括弧内の第 1 項は Θ が所与の場合には定数と見なすことができる。また, 右辺第 3 項についても $\text{tr}(\Theta_{gL} T \Delta T^T \Theta_{gL}^T) = \text{tr} \Delta = \text{tr} S$ となるから, やはり定数と見なすことができる。したがって, この場合の f を最小化する T は, $\text{tr}(\Delta \Theta_L^T \Theta_{gL} T)$ を最大化するような T であり, 因子スコアに注目した直交プロク拉斯テス回転を利用して

$$T = \Theta_{gL}^T \Theta_L \Delta (\Delta \Theta_L^T \Theta_{gL} \Theta_{gL}^T \Theta_L \Delta)^{-1/2}$$

と解析的に求めることができる。

4.3 計算機による実験

4.3.1 次元間の相関に関する実験

上記の等化法について、それぞれの性質を確認するために計算機を用いたシミュレーション実験を行った。実験の手順は以下のとおりである。

まず、 $N = 300$ 人の仮想的な受検者の $m = 2$ 次元の受検者特性値 $\theta_i (i = 1, 2, \dots, N)$ と $p = 40$ 個の問題項目の項目パラメータ \mathbf{a}_j および $d_j (j = 1, 2, \dots, p)$ を、 θ_i については、 $\theta_i \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$ 、 \mathbf{a}_j と d_j は $\mathbf{a}_j \sim N(\mathbf{0}, \sigma_a^2 I)$ 、 $\sigma_a^2 = 0.5^2$ 、 $d_j \sim N(0, 1)$ としてそれぞれ乱数で生成した。なお、受検者特性の次元間の相関の強さによる影響を考慮して、相関係数 ρ については 0.1, 0.3, 0.5, 0.7, 0.9 の 5 パターンを用意した。

つぎに θ_i と \mathbf{a}_j, d_j から再度正誤データを生成し、生成した正誤データについて項目パラメータおよび受検者の受検者特性値 θ_{gi} の推定を行った。

このようにして推定した受検者毎の受検者特性値を転置した θ_{gi}^T を受検者数分縦に並べた行列 Θ_g について、真の受検者特性値を転置した θ_i^T を受検者数分縦に並べた行列 Θ を等化先として等化を行った。等化の方法としては、上記の受検者特性値に注目した斜交プロクラステス回転法 (OQ)、重み付きプロクラステス回転法 (WP) に加えて、Hirtsch の方法 (HM)、直交プロクラステス回転法 (OT) の計 4 つの方法を用いた。

上記の手続きを相関係数 ρ の各値について、それぞれ $q=200$ 回繰り返し、つぎに示す $RMSE(\theta)$ 、 $RMSE(r)$ 、 $RMSE(A)$ の値を求めた。 $RMSE(\theta)$ は、

$$RMSE(\theta) = \sqrt{\frac{1}{mnq} \sum_{k=1}^q \|\Theta_{gk} W_{gk} - \Theta_k\|^2}$$

であり、値が小さいほど等化後の受検者特性値と等化先の受検者特性値が一致傾向にあることを表す。ただし、 Θ_{gk} は k 回目の繰り返しにおける等化対象の受検者特性値行列を、 Θ_k は同じく k 回目の繰り返しにおける等化先の受検者特性値行列を、 W_{gk} は等化係数行列をそれぞれ表す。

また、 $RMSE(r)$ は、等化後の θ_g の次元間の相関係数に関連する指標で、値が小さいほど等化後の θ_g の次元間の相関係数が等化先の θ の次元間の相関係数に平均的に近いことを表している。

$$RMSE(r) = \sqrt{\frac{1}{q} \sum_{k=1}^q \left[r(W_{gk}^T \theta_{gk}) - r(\theta_k) \right]^2} .$$

上式の $r(W_{gk}^T \theta_{gk})$ は、各繰り返し k における等化後の θ_g の次元間の相関係数、 $r(\theta)$ は等化先の θ の次元間の相関係数である。

$RMSE(\mathbf{a})$ は

$$RMSE(\mathbf{a}) = \sqrt{\frac{1}{mpq} \sum_{k=1}^q \|W_{gk}^{-1} A_{gk}^T - A_k^T\|^2}$$

で、 \mathbf{a} パラメータに関連する指標である。値が小さいほど等化後の \mathbf{a} パラメータが等化先の \mathbf{a} パラメータに近いことを表す。 A_{gk}^T は k 回目の繰り返しにおける等化対象の \mathbf{a} パラメータ行列を、 A_k^T は等化先の \mathbf{a} パラメータ行列を、それぞれ転置した行列を表す。

なお、シミュレーション実験のプログラムは R3.0.0 (R Development Core Team, 2013) によって作成した。また、多次元 IRT の項目パラメータと受検者の能力パラメータの推定には mirt パッケージ (Chalmers, 2012) を利用した。

$RMSE(\theta)$ に関する結果を図 (4.1) に、 $RMSE(r)$ に関する結果を図 (4.2) に、 $RMSE(\mathbf{a})$ に関する結果を表 (4.1) にそれぞれまとめる。

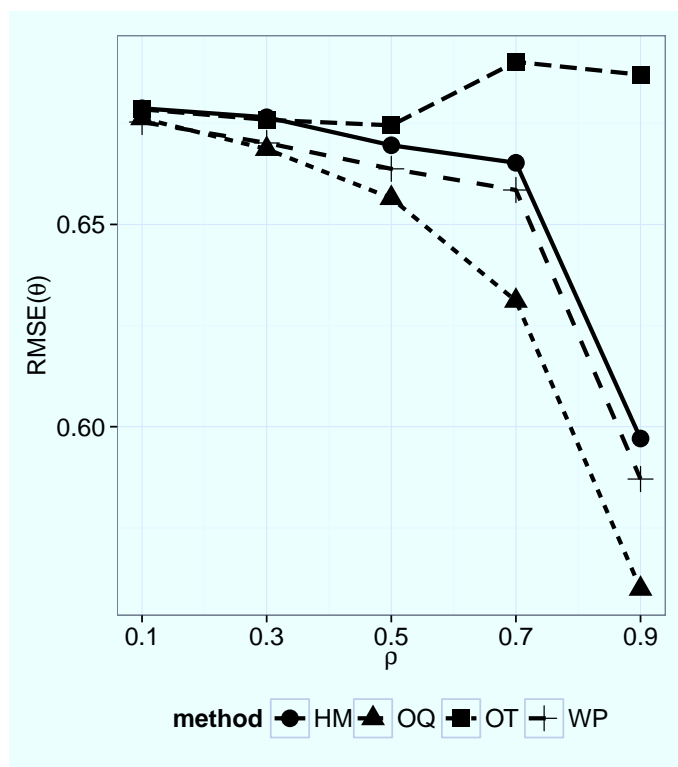


図 4.1: シミュレーション結果 $RMSE(\theta)$

まず、 $RMSE(\theta)$ についてみると、全般に次元間の相関係数が小さい場合には各手法間の差が小さいが、次元間の相関が大きくなるに従って等化法間の差が大きくなっていることがわかる。今回検討する等化法は、受検者特性値間

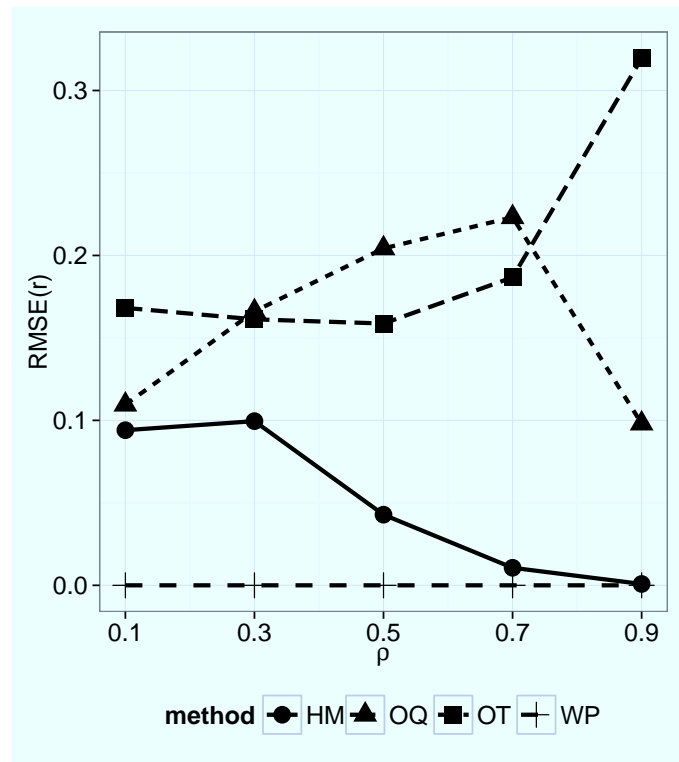


図 4.2: シミュレーション結果 $RMSE(r)$

の相関をどのように捉えるかという点に違いがあるため、相関が大きいほど差異が明確になるのは妥当な結果であると言える。

以下では、 $RMSE(\rho)$ 、 $RMSE(\alpha)$ の結果も含めて、4つの等化法に関する結果を順にみていく。まず直交プロクラステス回転法 (OT) では、特に ρ の値が大きく次元間に相関がある場合に、 $RMSE(\theta)$ 、 $RMSE(r)$ の値が双方とも大きくなっている。これは、因子スコアによる直交プロクラステス回転が、受検者特性値の次元間に直交性を仮定しており、受検者特性値の相関に上手く追従できないためであると考えられる。ただし、OT法で ρ の値が 0.9 と大きい場合に $RMSE(\alpha)$ の値が他の手法に比べて小さくなっている。OT法は、受検者特性値を等化先に近づけすぎない分、項目特性値の側から見ると安定した結果が得られている可能性もある。これは、特に多次元 IRT の等化において、等化対象のテストと等化先のテストで受検者の受検者特性値を近接させることが、必ずしも項目特性値を近接させることを同時に実現するものではないことを示す点で注目すべき結果であると言える。

つぎに、斜交プロクラステス回転法 (OQ) についてみると、いずれの ρ の値についても $RMSE(\theta)$ が他の手法に対して小さな値となっていた。これは、斜

表 4.1: ρ 毎の $RMSE(\mathbf{a})$

	OT	OQ	HM	WP
$\rho = 0.1$.73 (.21)	.73 (.21)	.73 (.21)	.73 (.21)
$\rho = 0.3$.72 (.21)	.73 (.21)	.71 (.21)	.72 (.21)
$\rho = 0.5$.74 (.20)	.78 (.21)	.74 (.20)	.74 (.20)
$\rho = 0.7$.77 (.21)	14.14 (13.95)	.79 (.20)	.79 (.21)
$\rho = 0.9$.78 (.21)	81.79 (79.22)	.96 (.26)	.96 (.27)

※括弧内の数値は各繰り返し毎の $RMSE(\mathbf{a})$ の標準偏差

交プロクラステス回転法が最も少ない制約の下で等化先と等化後の受検者特性値を近接させる構造となっていることの結果であると思われる。ただし、 ρ の値が大きい場合に $RMSE(r)$ の値が比較的大きくなっており、相関構造の再現性という点でみると斜交プロクラステス回転法は必ずしも優れた方法とはいえない点に注意が必要である。受検者特性値の次元間の相関関係はテストが測定する次元間の関係性に対応するものと考えられるが、等化によってこれが大きく変化してしまうことはテストの運用という面では望ましい性質ではない。

ところで ρ が大きい場合の斜交プロクラステス回転による等化後の θ_g の次元間相関を確認したところ、値が 1 に近いケースが見られた (図 4.3)。これは一部には実験において \mathbf{a} パラメータを小さく設定したことの結果であるが、斜交プロクラステス回転を利用した等化では等化後の受検者特性値の次元が縮退してしまう可能性があることを表すものである。テストが測定内容の面でも想定した次元数に対応しているとすれば、このような状況はやはり適切なものではない。なお、同様の現象は他の等化法、たとえば重み付きプロクラステス回転法においては見られなかった (図 4.4)。また、項目パラメータに関する指標である $RMSE(\mathbf{a})$ について見ても、斜交プロクラステス回転では次元間相関が 0.7 以上の場合に値が極めて大きくなっており、項目パラメータの再現についてもうまく行っていないことがわかる。特に次元間の相関が高い場合に斜交プロクラステス回転法を利用する際には、等化後のパラメータが妥当なものであるか十分注意する必要があると考えられる。

Hirsch の方法 (HM) と重み付きプロクラステス回転による方法 (WP) は、コンセプトにおいて類似性が高いため、比較する形で結果を確認する。まず図 (4.2) において、 ρ の値に関わらず重み付きプロクラステス回転による方法

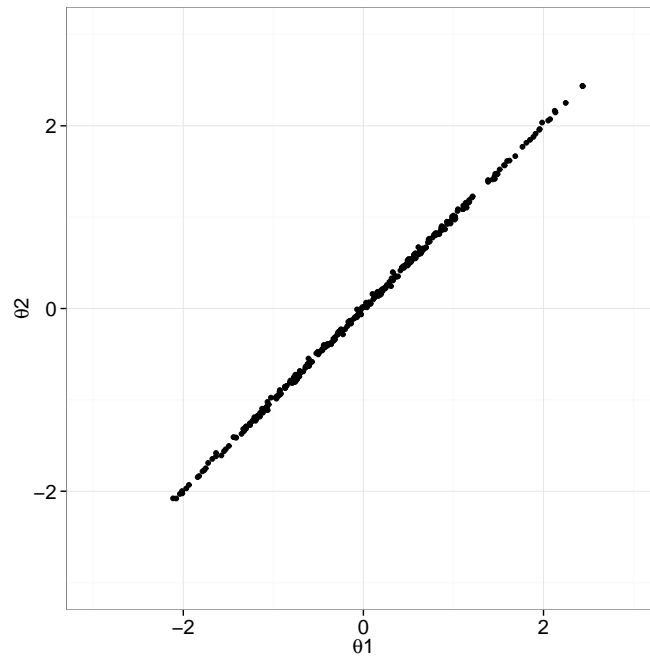


図 4.3: OQ 法による等化後の θ の例 $\rho = 0.9$ の場合

では $RMSE(r)$ の値が常にゼロになっていることがわかる。これは (4.9) 式で示したように、この手法では等化後の受検者特性値の次元間の相関が等化先の受検者特性値の次元間の相関に常に一致することの結果である。一方の Hirsch の方法では特に ρ の値が小さい場合に $RMSE(r)$ が比較的大きな値となっている。また、図 (4.1) の $RMSE(\theta)$ の結果を見ると、 ρ の値にかかわらず重み付きプロクラステス回転法のほうが値が小さくなっており、等化後の θ_g と等化先の θ の近接性という面でも重み付きプロクラステス法のほうが優れていることがわかる。一方、 $RMSE(\mathbf{a})$ については、Hirsch の方法と重み付きプロクラステス回転法の値はほぼ同じ水準となっているが、 $\rho = 0.3$ と相関が小さい場合にはわずかながら Hirsch の方法のほうが値が小さくなっている。したがって、共通受検者による等化において次元間の相関の再現性が求められる場合には、Hirsch の方法よりも重み付きプロクラステス回転法による方法を利用するほうが、一般に良好な結果が得られるものといえる。

これらの結果から、受検者特性値間の相関が 0.3 を目安として一定の水準を超えるような場合の共通受検者による等化法としては、重み付きプロクラステス回転による方法 (WP) を利用することが適切だと考えられる。一方、特に等化対象のテストと等化先のテストで受検者特性値を一致させる要請が強い場合には、斜交プロクラステス回転法 (OQ 法) の利用が検討されてもよい。ただし、この場合には受検者特性値間の相関構造は再現されないこと、受検者

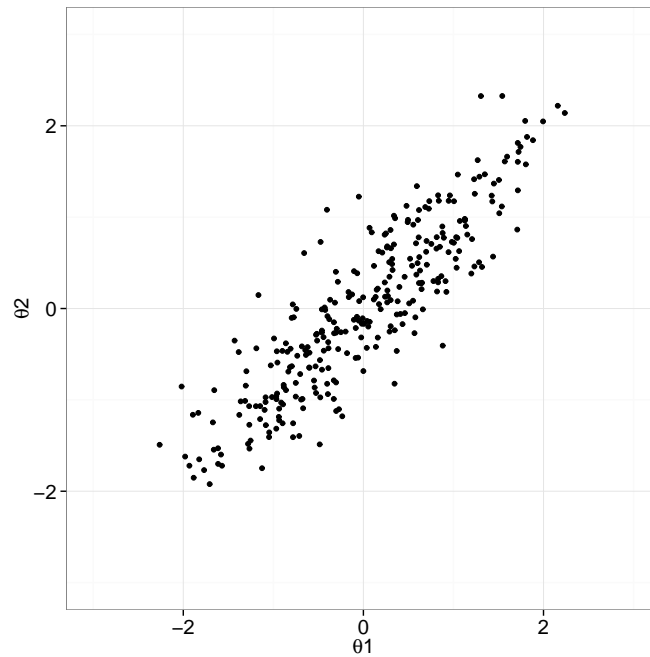


図 4.4: WP 法による等化後の θ の例 $\rho = 0.9$ の場合

特性値間の相関が強い場合には項目パラメータや受検者特性値そのものが不安定になることに注意する必要がある。

4.3.2 項目数に関する実験

つぎに、問題項目数を変化させた場合の各等化法の違いを確認するために、受検者特性値間の相関係数は $\rho = 0.7$ に固定し、項目数 p について 20, 40, 100, 200 の 4 パターンを用意して実験を行った。相関係数と項目数以外の設定および評価の指標は相関係数に関する実験と同様である。 $RMSE(\theta)$ に関する結果を図 (4.5) に、 $RMSE(r)$ に関する結果を図 (4.6) に、 $RMSE(\mathbf{a})$ に関する結果を表 (4.2) にそれぞれまとめる。

まず、 $RMSE(\theta)$ の結果をみると、項目数が増えるほどどの手法でも明らかに値が小さくなる傾向があることがわかる。これは、項目数を増やすことで等化前の受検者特性値の推定精度が高くなることの効果であると考えられる。また、 $RMSE(\theta)$ の値は、項目数に関わらず OQ 法, WP 法, HM 法, OT 法の順に小さくなっており、等化対象のテストと等化先のテストの受検者特性値の一致度合という観点でみると、この順で有効な方法であると考えられる。

$RMSE(\rho)$, $RMSE(\mathbf{a})$ の結果も含めて 4 つの等化法に関する結果を順にみていくと、直交プロクラステス回転法 (OT) では $RMSR(\theta)$, $RMSR(r)$ の

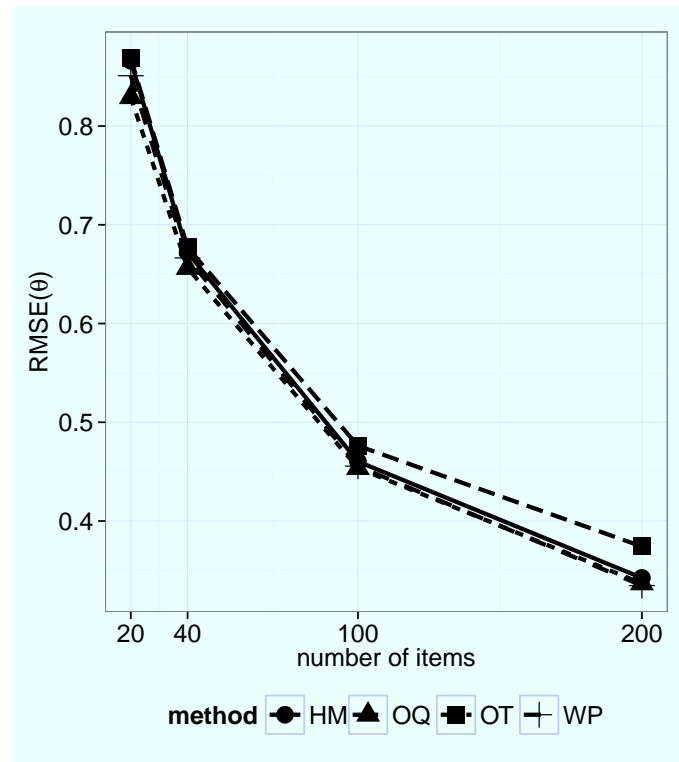


図 4.5: シミュレーション結果 $RMSE(\theta)$

両指標ともに他の等化法と比較して相対的に大きな値となっており、特に項目数が多い場合にこの傾向が顕著であることがわかる。OT法は今回検討する4つの手法の中で最も制約の強い方法であり、受検者特性値の近接度、相関構造の再現性の面では他の手法のほうが有利であると言える。 $RMSR(\theta)$ 、 $RMSR(r)$ の値が大きくなるのは、このことを反映したものである。ただし $RMSE(a)$ をみると、いずれの項目数の場合についても他の手法よりもOT法の値が小さくなっており、相関係数の実験結果と併せて考えると、特に次元間の相関が強い場合には、項目特性値については直交プロクラステス回転法が安定した結果をもたらすものと考えられる。

斜交プロクラステス回転法(OQ)は、 $RMSR(\theta)$ が項目数に関わらず4つの等化法の中で最も小さい値となっており、等化対象のテストと等化先のテストの受検者特性値を一致に近づけるという点で優れた方法であるといえる。前述の項目数を40に固定した相関係数に関する実験では、受検者特性値の次元間の相関が高い場合に相関構造や項目パラメータの再現性において問題が見られた。一方、項目数に関する結果をみると、問題数が多くなるほど $RMSE(r)$ 、 $RMSE(a)$ の値は改善し、特に問題数が80問を超えると $RMSE(a)$ の値は

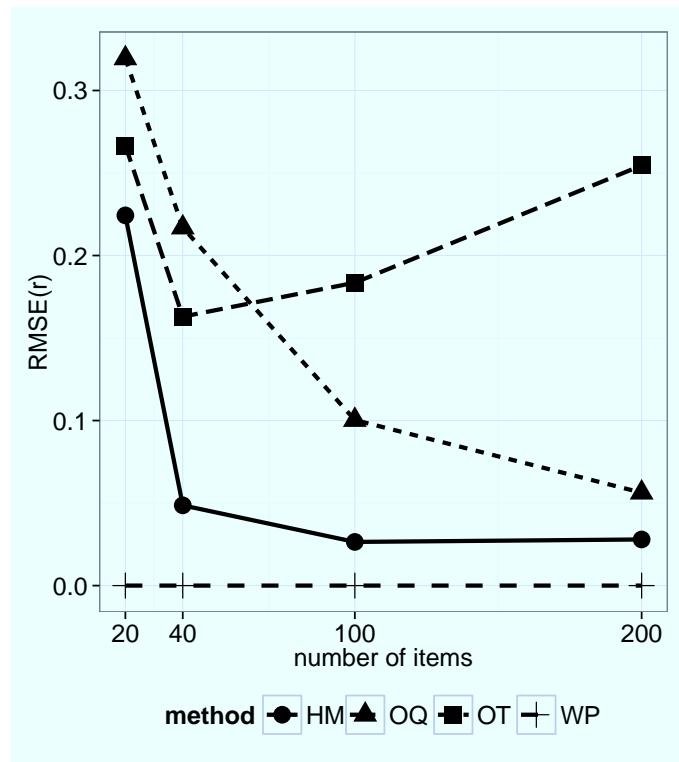


図 4.6: シミュレーション結果 $RMSE(r)$

他の等化法と遜色ない水準となっている。これらの結果から、相関構造の再現性があまり重要ではなく受検者特性値の近接性が重視される場合で、かつテストを構成する項目数が十分多いような状況では、斜交プロクラステス回転は有効な方法であると考えられる。

Hirsch の方法 (HM) と重み付きプロクラステス回転による方法 (WP) をみると、 $RMSR(\theta)$, $RMSR(r)$, $RMSR(A)$ の各値とも、項目数が多くなるにしたがって両手法での差は小さくなっており、特に項目数が 80 以上の場合にはかなり近接した結果となっていることがわかる。ここから、Hirsch の方法と重み付きプロクラステス回転による方法は、特に項目数が多い場合には類似した特徴を示すものと考えられる。ただし、重み付きプロクラステス回転による方法では $RMSR(r)$ の値が常にゼロとなることを考慮すれば、やはり重み付きプロクラステス回転法のほうが実用上は有利であると思われる。

表 4.2: 項目数毎の $RMSE(\mathbf{a})$

項目数	OT	OQ	HM	WP
$p = 20$.77 (.26)	2.24 (1.85)	.77 (.26)	.81 (.30)
$p = 40$.76 (.21)	.85 (.28)	.77 (.21)	.77 (.21)
$p = 100$.84 (.25)	.88 (.25)	.85 (.25)	.85 (.25)
$p = 200$.90 (.27)	.93 (.27)	.92 (.27)	.92 (.27)

※括弧内の数値は各繰り返し毎の $RMSE(\mathbf{a})$ の標準偏差

4.4 まとめと課題

実験の結果, 等化後の受検者特性値と等化先の受検者特性値を近接させることを第一に考える場合には, 斜交プロクラステス回転法が最も有効な手法であることが示された. ただし, テストを構成する項目数が少ない場合, 斜交プロクラステス回転法では等化後の受検者特性値の次元間の相関係数が等化先のそれを再現したものとならない傾向があることが把握できた. 受検者特性値の次元間の相関は, テストが測定している受検者特性間の関係性を表すものであるが, 斜交プロクラステス回転を利用した等化では, この関係が変化してしまう可能性がある点には注意が必要である.

また, 等化後の受検者特性値の次元間の相関を等化先に一致させることが重視される場合には, 重み付きプロクラステス回転による方法が有効であることが確認できた. 重み付きプロクラステス回転では, 等化後の受検者特性値についても斜交プロクラステス回転法に次いで等化先の受検者特性値との近接性が高い結果となっており, 特に受検者特性値の次元間の相関が小さい場合には斜交プロクラステス回転法との差が小さい傾向であった. したがって, 等化にあたって受検者特性値の次元間の相関を等化先と近づけることにも配慮する必要がある場合, 重み付きプロクラステス回転による方法の総合的な有用性は高いものと考えられる.

ところで, 本章の実験の結果から, 等化後の受検者特性値を等化先に近接させるという面で有効な等化法が, 必ずしも共通項目のパラメータを最も近接させる等化法とはならないことが示唆された. 次章では, この問題について改めて取り上げ, 共通受検者と共通項目の情報を同時に考慮した等化法について検討する.

第5章 共通受検者と共通項目を同
時に利用した
多次元IRTの等化法

5.1 問題

前章の実験結果から、共通受検者の特性値を等化先に近づけることを強く考慮した等化法の場合、共通項目については却って等化先と等化後の値の乖離が大きくなる傾向があり、等化後の受検者特性値を等化先に近接させるという面で有効な等化法が、必ずしも共通項目のパラメータを最も近接させるというわけではないことがわかった。また、一般に、共通項目による等化と共通受検者による等化では異なる基準が用いられることから、両方の等化デザインが利用できる場合には、どちらを利用するかによって得られる解は異なったものとなると考えられる。

ところで、IRTにおける等化は、通常複数のテストのスコアを同一の尺度上で比較可能とするための操作であるが、特に多次元 IRT においては、つぎのような場合にも形式的に等化の手続きが活用できる。まず、それまで1次元の IRT を利用して運用されてきた複数のテストを多次元 IRT によって統合する場合である。このような場面では、各テストについて従来のスコアとの一貫性のある程度確保した上で、多次元 IRT によるスコアを改めて算出することとなるが、このために1次元の IRT によるスコアもしくは項目パラメータをターゲットとして多次元 IRT の各パラメータを等化する手続きを利用することが考えられる。また、多次元 IRT の適用を前提として新規に開発されたテストにおいても、測定内容の安定性を確保する観点から、一旦1次元の IRT を利用して想定される各測定次元に対応する項目パラメータや受検者特性値を算出し、しかる後にこれをターゲットとして多次元 IRT の項目パラメータを等化する手続きを考えることができる。これらの場合にはいずれも、共通項目による等化と共通受検者による等化の両方が利用可能である。

上記のように、共通項目と共通受検者が共に利用可能な場合には、両方の情報を併せて考慮することで、バランスのよい等化が実現できる可能性があるが、このような手法は現状では提案されていない。

本章では、共通項目と共通受検者の両方の等化法が利用可能な場面に注目し、これら2つの方法を統合した新たな等化法を提案する。

5.2 共通項目と共通受検者の両方を考慮した等化法

1次元の複数の尺度を多次元 IRT の枠組みで統合する場合、受検者特性値の次元間の相関は測定内容の特徴を反映した重要な特性であると考えられる。また、共通受検者の特性値の測定次元毎の分散については、等化先と等化後で同じと考えるのが自然である。ここでは、等化後の受検者特性値の分散共分散行列が等化先と一致することを制約として考え、この制約下で共通受検者と共通項目の双方を同時に利用した等化法について検討する。

さて、前章で示した共通受検者による重み付きプロクラステス回転による方法は、等化後の受検者特性値の分散共分散行列を等化先の分散共分散行列

と一致させるという制約条件下で共通受検者デザインの場合の等化を実現するものであった。以下ではまず、同様の制約条件下で共通項目に注目した等化を行う手順について考える。つぎに、受検者特性値による基準と共通項目による基準を統合した等化基準を設定することで、共通受検者と共通項目の両方を同時に考慮した等化法を構成する。

5.2.1 共通項目に注目した重み付きプロクラステス回転による方法

いま、等化係数行列 W_g が前章でみた共通受検者による重み付きプロクラステス回転法と同様、次式の形で表されるものとする。

$$W_g = P\Lambda^{-1/2}T\Delta^{1/2}Q^T \quad (5.1)$$

ただし、 Λ , P はそれぞれ等化対象の受検者特性値 θ_g の分散共分散行列 S_g をスペクトル分解した場合の固有値を対角に並べた行列と、対応する固有ベクトルを順に並べた行列、 Δ , Q は等化対象の受検者特性値 θ の分散共分散行列 S をスペクトル分解した場合の固有値を対角に並べた行列と固有ベクトルを順に並べた行列であり、 $S_g = P\Lambda P^T$, $S = Q\Delta Q^T$, $PP^T = P^T P = Q^T Q = QQ^T = I$ である。前章で確認したとおり、このときには等化後の受検者特性値の分散共分散行列は等化先のものと同様に一致する。

ここで、共通項目の等化後の \mathbf{a} パラメータ行列と等化先の \mathbf{a} パラメータ行列の距離を次のように定義する。

$$f_r = \frac{1}{2mp} \text{tr}[(A - A_g(W_g^{-1})^T)^T (A - A_g(W_g^{-1})^T)] \quad (5.2)$$

上式の $(W_g^{-1})^T$ は (5.1) 式より

$$(W_g^{-1})^T = P\Lambda^{1/2}T\Delta^{-1/2}Q^T$$

となる。したがって、(5.2) 式は

$$\begin{aligned} & \frac{1}{2mp} \text{tr}[(A - A_g(W_g^{-1})^T)^T (A - A_g(W_g^{-1})^T)] \\ &= \frac{1}{2mp} \text{tr}[(A - A_g(W_g^{-1})^T)(A - A_g(W_g^{-1})^T)^T] \\ &= \frac{1}{2mp} \text{tr}[(AQ\Delta^{1/2} - A_gP\Lambda^{1/2}T)(\Delta^{-1/2}Q^T) \\ & \quad (\Delta^{-1/2}Q^T)^T (AQ\Delta^{1/2} - A_gP\Lambda^{1/2}T)^T] \\ &= \frac{1}{2mp} \text{tr}[(A_L - A_{gL}T)\Delta^{-1}(A_L - A_{gL}T)^T] \end{aligned} \quad (5.3)$$

と書ける。ただし、 $A_L = AQ\Delta^{1/2}$, $A_{gL} = A_gP\Lambda^{1/2}$ とした。

(5.3) 式を最小化する直交回転行列 T を求めることで、等化先と等化後の受検者特性値の分散共分散行列を一致させるという制約下で共通項目による等化を実現することができる。直交回転行列 T は、たとえば因子分析の場合の重み付きプロクラステス回転法 (Koschat & Swayne, 1991), Gradient Projection 法 (Jennrich, 2001) などを利用して数値的に求めることができる。なお、この方法では主に共通項目に関連する基準を用いて等化を行っているが、制約として等化先のテストの受検者特性値の分散共分散行列を利用するため、この値を把握しておく必要がある点に注意する。各テストフォームについて共通受検者個々のスコアが存在する場合には、無論この条件は満たされる。

5.2.2 共通項目と共通受検者の両方を考慮した重み付きプロクラステス回転による方法

ここで、共通項目と共通受検者に関する等化基準を統合することを考えるが、これに先立って先行研究の方法を含めて本章で取り上げる等化法について整理しておく。表 (5.1) は、ここまでに確認した先行研究の方法と本章で提案する等化法の特徴をまとめたものである。等化法は大きく、共通項目によるものと共通受検者によるものに分けることができる。また、それぞれの等化法を受検者特性値の分散共分散行列に制約のあるものと制約のないものに分類することができる。Min の方法、Hirsch の方法はそれぞれ、共通項目と共通受検者による代表的な等化法であり、共に分散共分散行列に制約がない場合の等化に対応した手法である。

表 5.1: 本章で取り上げる多次元 IRT の等化手法

	共通項目による等化法	共通受検者による等化法
分散共分散制約なし	Min(2003)の方法	Hirsch(1989)の方法
分散共分散制約あり	共通項目による WP 法 (本研究 5.2.1 章)	共通受検者による WP 法 (本研究 4.2.2 章)
	共通項目と共通受検受検者の両方を考慮した WP 法 (提案手法/本研究 5.2.2 章)	

共通項目に注目した重みつきプロクラステス回転による方法は、分散共分散行列に制約がある場合の共通項目による等化にあたる。以降では、この方法と前章で示した分散共分散行列に制約のある共通受検者による重みつきプロクラステス回転法の2つを合成することで、共通項目と共通受検者の基準を統合し双方の情報を利用した等化法を構成する。

さて、2つの等化基準のうち、共通項目に関するものについては、(5.3)式で表された。もう一方の受検者特性値に注目した等化の基準については(4.5)式を利用することができる。ただし、受検者数の影響を除くため、共通項目の場合と揃えて受検者数と次元数でこれを除しておく。共通受検者の場合の等化基準を f_s と書くことにすると、

$$f_s = \frac{1}{2mn} \text{tr} [(\Theta_L - \Theta_{gL}T)\Delta(\Theta_L - \Theta_{gL}T)^T] \quad (5.4)$$

となる¹。

等化のためには、共通項目による基準と共通受検者による基準を同時に最小化するような等化係数行列を求めることが課題となるが、これは一種の多目的最適化問題とみなせる。ここでは、共通項目による基準と共通受検者による基準に関する重み λ ($0 \leq \lambda \leq 1$) を導入し、 f_r と f_s の重み付き和として両基準を統合し、これを最小化することを考える。この最小化問題は、

$$\min \quad f = \lambda f_r + (1 - \lambda) f_s \quad (5.5)$$

$$\text{s.t.} \quad T^T T = I \quad (5.6)$$

と書ける。 λ は共通受検者と共通項目のどちらを重視するかを表す重みであり、 $\lambda = 0$ の場合には、前章の共通受検者を利用した等化に相当する。また $\lambda = 1$ の場合には共通項目に注目した重み付きプロクラステス回転法による等化が実現できる。したがって、提案手法はこれらの方法を一般化した手法であるとみなすことができる。

なお、(5.5)式の f の T による微分は、

$$\begin{aligned} \frac{\partial f}{\partial T} &= \lambda \frac{\partial f_r}{\partial T} + (1 - \lambda) \frac{\partial f_s}{\partial T} \\ &= \frac{\lambda}{mp} (A_L^T A_{gL} + A_{gL}^T A_{gL} T) \Delta^{-1} \\ &\quad + \frac{1 - \lambda}{mn} (\Theta_L^T \Theta_{gL} + \Theta_{gL}^T \Theta_{gL} T) \Delta \end{aligned}$$

となる。目的関数の微分が明示的に書けるため、 f を最小化する直交回転行列 T は Gradient Projection 法 (Jennrich, 2001) によって求めることができる。また、次元数が3以上の場合には因子分析の場合の逐次平面回転 (柳井 他、

¹(5.4)式は $(2mn)^{-1} \text{tr}(\Theta_L \Delta \Theta_L^T - 2\Delta \Theta_L \Theta_{gL} T) + 1/2$ と展開することもできる。この場合、一部の項が定数となるため計算機での演算などで有利なことがあるが、本文では(5.3)式の対応が把握しやすい表現を採用した。

1990) のような方法を利用することができる。 T は回転と次元の入れ替えを実現する行列であり、次元の入れ替えはこの場合、行列 A と A_g 、 Θ と Θ_g の列の並び順を一致させる操作を意味する。

ところで、逐次平面回転法は2つの次元の組に注目して平面上での回転を順次繰り返す方法であるが、逐次平面回転における平面上での回転には軸の入れ替えの操作が含まれない²ため、この方法では次元の並び替えを実現することができない。また、Gradient Projection 法の場合にも、等化対象と等化先のパラメータ行列の列の並び順が対応していない場合には目的関数が複雑な形状となるため局所解が生じやすくなる。したがって実際の計算において、次元の並び順があらかじめ明らかである場合には、等化対象と等化先でこれが一致するように事前にパラメータ行列の列を並び替えておく必要がある。また、後述する計算機実験のように次元の並び順を事前に揃えることが難しい場合には、等化対象のパラメータ行列の可能な列の並び替えについて目的関数の値をそれぞれ計算し、最も小さい値を与える結果を採用する。なお、Min の方法、Hirsch の方法、共通受検者による重み付きプロクラステス回転法では解析的に求めた等化係数行列に次元の入れ替えの操作が含まれており、この操作を取り出して考える必要はない。

なお、これまでの議論から、提案手法はつぎの2つの特徴をもつものと考えられる。第1に、等化先と等化後の受検者特性値の次元間の分散および相関係数行列が必ず一致する。第2に、事前に与える重み λ によって、共通項目を重視した等化を行うか共通受検者に注目した等化を行うのかを調整することが可能である。

5.3 計算機による実験

提案手法 (WP) の特性を確認するために計算機による実験を行った。多次元 IRT の等化には等化法の他に、次元数、受検者特性値の次元間の相関係数の水準と構造、項目数、受検者数、 α パラメータの構造、パラメータの推定アルゴリズムなど、1次元の IRT の場合に比較して非常に多様な要素が相互に組み合わさる形で影響すると考えられる。ここでは、これらの要因のうち、受検者特性値の次元間の相関の大きさと次元の数に注目し、それぞれの影響について検討した。この際、既存の手法として、Min の方法、Hirsch の方法を取り上げ、提案手法と共に計算を行い結果を比較した。なお、共通受検者による重み付きプロクラステス回転法は、提案手法で $\lambda = 0.0$ とした場合と同一の手法となるため、提案手法に包含されるものとして扱った。各実験の設定と結果について、順に確認していく。

²因子分析の文脈では因子の回転、軸の入れ替え、伸縮等の操作を総称して「回転」と呼ぶことが一般的になっているが、逐次平面回転法における回転は語の本来の意味における回転操作に対応するものであり、次元の入れ替え等の操作は通常、逐次平面回転法のアルゴリズムには含まれない。

5.3.1 次元間の相関係数に関する実験

まず, 受検者特性値の次元間の相関係数の提案手法に対する影響を調べるために, つぎの実験を行った. 受検者特性値の次元数を $m = 2$, 項目数を $p = 50$ とし, \mathbf{a} パラメータは各次元について値の範囲が 0.1 から 0.7 の一様乱数を用いて生成した. d パラメータについては, (2.14) 式の関係から, まず平均 0.0, 分散 1.0 の正規乱数によって b_j を生成し, これを各項目について \mathbf{a} のノルムを乗じて符号を反転した値とした. 受検者特性値 Θ は, 平均 0.0, 分散 1.0, 相関 ρ の 2次元正規分布から 500 人分を乱数で生成した値を各次元について平均 0 に中心化したものとした. 次元間の相関 ρ については, 0.0, 0.3, 0.6, 0.9 の 4 つのパターンを用意した. このようにして生成した項目パラメータと受検者特性値について, (2.11) 式の多次元 IRT モデルに従って回答データ行列を作成し, あらためて項目パラメータと受検者特性値を推定した. この際, 等化前の \mathbf{a} パラメータ行列の推定値は promax 解とした. 受検者特性値については, 独立な標準正規分布を事前分布としてベイズ推定により事後分布を求め, 事後期待値を平均 0 に中心化したものを推定値として採用した.

つぎに, 推定した \mathbf{a} パラメータ行列を A_g , 受検者特性値行列を Θ_g とし, 元の項目パラメータ \mathbf{a}_j^T を縦に並べた行列 A および受検者特性値行列 Θ を等化先として提案手法によって等化を行った. 提案手法の重み λ は 1.0, 0.75, 0.5, 0.25, 0.0 の 5 パターンとし, 相関係数 ρ の各パターンについて上記の計算を $q = 100$ 回ずつ繰り返した. 計算プログラムには R3.2.3 (R Development Core Team, 2015) と mirt パッケージ (Chalmers, 2012) を利用した.

等化の結果について, 既存の等化法を含むそれぞれの手法毎につぎの $RMSE(\boldsymbol{\theta})$, $RMSE(\mathbf{a})$, $RMSE(S)$ の 3 つの指標の値を確認した. まず, $RMSE(\boldsymbol{\theta})$ は,

$$RMSE(\boldsymbol{\theta}) = \sqrt{\frac{1}{mnq} \sum_{k=1}^q \|\Theta_{gk} W_{gk} - \Theta_k\|^2}$$

であり, 値が小さいほど等化後の受検者特性値と等化先の受検者特性値が一致傾向にあることを表す. ただし, Θ_{gk} は k 回目の繰り返しにおける等化対象の受検者特性値行列を, Θ_k は同じく k 回目の繰り返しにおける等化先の受検者特性値行列を, W_{gk} は等化係数行列をそれぞれ表す. つぎに $RMSE(\mathbf{a})$ は,

$$RMSE(\mathbf{a}) = \sqrt{\frac{1}{mpq} \sum_{k=1}^q \|W_{gk}^{-1} A_{gk}^T - A_k^T\|^2}$$

で, \mathbf{a} パラメータに関連する指標である. 値が小さいほど等化後の \mathbf{a} パラメータが等化先の \mathbf{a} パラメータに近いことを表す. A_{gk}^T は k 回目の繰り返しにおける等化対象の \mathbf{a} パラメータ行列を, A_k^T は等化先の \mathbf{a} パラメータ行列を, そ

れぞれ転置した行列を表す. 3つ目の $RMSE(S)$ は,

$$RMSE(S) = \sqrt{\frac{1}{qm^2} \sum_{k=1}^q \|V(\Theta_{gk}W_{gk}) - V(\Theta_k)\|^2}$$

である. ただし, 上式の $V(\Theta_{gk}W_{gk})$ は k 回目の繰り返しにおける等化後の θ_g の分散共分散行列を, $V(\Theta_k)$ は k 回目の繰り返しにおける等化先の θ の分散共分散行列を表す. $RMSE(S)$ の値が小さいほど等化後の受検者特性値の分散共分散行列が等化先に近いことを表す. なお, $RMSE(\mathbf{a})$, $RMSE(\theta)$ と同様, $RMSE(S)$ は明らかに常に非負であり, 値が 0 になるのは全ての繰り返し (k) において等化先と等化後の受検者特性値の分散共分散行列が一致している場合である.

$RMSE(\theta)$ について表 (5.2) に, $RMSE(\mathbf{a})$ について表 (5.3) に, $RMSE(S)$ の結果について表 (5.4) にそれぞれまとめる. また, 各繰り返し (k) 毎に $RMSE(\theta)$ および $RMSE(\mathbf{a})$ を計算した値を, それぞれ $RMSE(\theta)_k$, $RMSE(\mathbf{a})_k$ とし, 100 回の繰り返し全体における 25 パーセンタイル点 (Q1), 中央値, 75 パーセンタイル点 (Q3) を併せて表に記載した.

表 5.2: 次元間の相関係数 ρ を変化させた場合の各手法の $RMSE(\theta)$ の値

		WP						
		$\lambda = 1.0$	$\lambda = .75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.0$	Hirsch	
$\rho = .0$	$RMSE(\theta)$.976	.965	.963	.962	.962	1.029	.952
	Q1	.937	.914	.908	.906	.906	.939	.898
	中央値	.975	.957	.956	.956	.956	1.014	.946
	Q3	1.015	1.011	1.011	1.011	1.011	1.091	1.005
$\rho = .3$	$RMSE(\theta)$.873	.863	.860	.859	.859	1.118	.859
	Q1	.843	.834	.831	.830	.830	.997	.831
	中央値	.877	.863	.860	.859	.859	1.087	.859
	Q3	.901	.896	.890	.888	.888	1.219	.889
$\rho = .6$	$RMSE(\theta)$.720	.713	.710	.708	.708	1.144	.710
	Q1	.701	.696	.695	.694	.694	1.012	.696
	中央値	.719	.715	.711	.709	.709	1.118	.711
	Q3	.734	.725	.724	.723	.723	1.210	.723
$\rho = .9$	$RMSE(\theta)$.508	.495	.488	.485	.485	1.218	.488
	Q1	.487	.482	.477	.475	.474	1.065	.479
	中央値	.502	.494	.490	.486	.486	1.175	.487
	Q3	.518	.509	.499	.494	.493	1.348	.498

表 5.3: 次元間の相関係数 ρ を変化させた場合の各手法の $RMSE(\mathbf{a})$ の値

		WP						
		$\lambda = 1.0$	$\lambda = .75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.0$	Min	Hirsch
$\rho = .0$	$RMSE(\mathbf{a})$.205	.210	.216	.221	.224	.190	.221
	Q1	.188	.193	.195	.199	.202	.174	.193
	中央値	.203	.208	.213	.217	.219	.188	.216
	Q3	.217	.223	.234	.239	.242	.206	.241
$\rho = .3$	$RMSE(\mathbf{a})$.234	.237	.244	.251	.257	.187	.257
	Q1	.217	.219	.224	.227	.232	.175	.232
	中央値	.231	.233	.238	.247	.254	.187	.253
	Q3	.244	.247	.256	.269	.277	.199	.275
$\rho = .6$	$RMSE(\mathbf{a})$.257	.257	.263	.271	.280	.181	.289
	Q1	.229	.230	.236	.244	.253	.170	.258
	中央値	.252	.250	.257	.264	.273	.182	.284
	Q3	.284	.282	.286	.291	.300	.190	.311
$\rho = .9$	$RMSE(\mathbf{a})$.368	.364	.384	.415	.452	.179	.488
	Q1	.332	.332	.344	.354	.370	.172	.398
	中央値	.356	.354	.370	.393	.425	.179	.479
	Q3	.385	.385	.403	.436	.474	.187	.525

表 5.4: 次元間の相関係数 ρ を変化させた場合の各手法の $RMSE(S)$ の値

	WP					Min	Hirsch
	$\lambda = 1.0$	$\lambda = .75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.0$		
$\rho = .0$.000	.000	.000	.000	.000	.816	.053
$\rho = .3$.000	.000	.000	.000	.000	1.225	.033
$\rho = .6$.000	.000	.000	.000	.000	1.681	.033
$\rho = .9$.000	.000	.000	.000	.000	2.152	.022

まず、表 (5.2) の $RMSE(\boldsymbol{\theta})$ について、提案手法 (WP) に関する結果をみると、全般には相関係数 ρ の値によらず、重み λ の値が 0.0 に近いほど $RMSE(\boldsymbol{\theta})$ の値が小さくなっていることがわかる。また表 (5.3) からは、 λ が 1.0 に近いほど $RMSE(\mathbf{a})$ の値が全般に小さくなっていることがわかる。提案手法は重み λ によって共通項目と共通受検者のいずれを重視するかを調整することを目指すものであるが、上記の結果はこの特徴によるものである。

つぎに、表 4 の $RMSE(S)$ の結果についてみると、提案手法では重み λ 、相関 ρ の水準によらず、値がすべて 0.000 となっている。提案手法は等化先と等化後の受検者特性値の分散と相関が一致することを制約とした手法であり、この結果は提案手法の理論上の特徴に合致するものである。

さて、表 (5.2) の結果をあらためて詳細にみると、たとえば $\rho = 0.0$ の場合、重み $\lambda = 0.25$ と $\lambda = 0.0$ で $RMSE(\boldsymbol{\theta})$ は 0.962 と同じ値になっている。 ρ が他の値の場合にも結果は同様である³。また表 (5.3) では、 $\rho = 0.6$ の場合に、 $\lambda = 1.0$, $\lambda = 0.75$ で $RMSE(\mathbf{a})$ がいずれも 0.257 と同じ値となっており、 $\rho = 0.9$ の場合には、 $\lambda = 0.75$ では 0.364 $\lambda = 1.0$ のときには 0.368 と、僅かではあるが λ が大きいほうが $RMSE(\mathbf{a})$ の値が大きくなっている。 ρ の値ごとの繰り返し (k) の結果について値を詳しく確認したところ、100 回の繰り返しそれぞれの λ の水準に関する $RMSE(\mathbf{a})_k$ の差 400 箇所 (= 100 × 4) のうち、 $\rho = 0.0$ の場合には 9 箇所、 $\rho = 0.3$ では 36 箇所、 $\rho = 0.6$ では 76 箇所、 $\rho = 0.9$ では 82 箇所値の大小関係が λ の値から想定されるものと逆になっていた。同様に、 $RMSE(\boldsymbol{\theta})_k$ では $\rho = 0.0$ の場合 68 箇所、 $\rho = 0.3$ で 61 箇所、 $\rho = 0.6$ で 63 箇所、 $\rho = 0.9$ で 48 箇所値の大小関係が λ の値から想定されるものと逆になっていた。

これらの結果から、提案手法において重み λ によって共通受検者に関する基準と共通項目に関する基準を調整する機能は多くの場合には想定どおり機能しているが、調整は完全なものではなく、 \mathbf{a} パラメータについては次元間の相関が大きい場合に、受検者特性値については次元間の相関が小さい場合に、重み λ の大小関係と想定される結果の大小関係が逆転する現象が生じやすい傾向があることがわかる。ただし実際の活用場面では、たとえば共通項目を重視したい場合について、重み λ を 1.0 としたときよりも、 λ をより小さい値に設定した際に $RMSE(\mathbf{a})$ が小さくなるのであれば、その結果を採用すればよく、重みによる調整が完全ではないことの応用上の影響は大きくはないものと考えられる。

ところで、表 (5.2)、表 (5.3) から提案手法において $RMSE(\boldsymbol{\theta})$ 、 $RMSE(\mathbf{a})$ は λ に対して線形の関係ではなく、 $RMSE(\boldsymbol{\theta})$ については λ が 0.0 に近い領域で、 $RMSE(\mathbf{a})$ については λ が 1.0 に近い領域で傾きが緩やかになる傾向があることがわかる。このことは提案手法を利用する際、たとえば $\lambda = 0.0$ では $RMSE(\mathbf{a})$ が大きすぎる場合、 λ を 0.25 などやや大きな値に設定すること

³ただし、小数点以下第 4 位まで値を確認すると、いずれの ρ においても $\lambda = 0.25$ のときよりも $\lambda = 0.0$ の場合のほうが、 $RMSE(\boldsymbol{\theta})$ は小さくなっていた。

で $RMSE(\boldsymbol{\theta})$ の値をあまり悪化させずに $RMSE(\mathbf{a})$ の値を改善できることを意味するものである。実務上は、このような操作によって総合的な観点から適切な λ を選ぶような手続きが有効であるものと思われる。

ここで、共通受検者に基づく等化法である Hirsch の方法と提案手法の結果を比較する。提案手法では重み $\lambda = 0.0$ とした場合に共通受検者に注目した等化が実現されるため、この場合について結果を比べると、まず表 (5.2) から ρ が 0.0 の場合には、提案手法の $RMSE(\boldsymbol{\theta})$ は 0.962 と Hirsch の方法の 0.952 よりも大きな値となっている。 $\rho = 0.3$ では提案手法、Hirsch の方法ともに $RMSE(\boldsymbol{\theta}) = 0.859$ で同じ値である。一方、 ρ が 0.6, 0.9 と受検者特性値の次元間の相関が大きい場合には、提案手法のほうが Hirsch の方法よりも $RMSE(\boldsymbol{\theta})$ の値がやや小さくなっていることがわかる。また、表 (5.3) から $RMSE(\mathbf{a})$ についてみると、 $\rho = 0.0$ の場合には提案手法で 0.224, Hirsch の方法は 0.221 と提案手法のほうが値が大きい。 $\rho = 0.3$ の場合には、 $RMSE(\mathbf{a}) = 0.257$ で同じ値である。一方、 ρ が 0.6, 0.9 と大きい場合については提案手法のほうが Hirsch の方法よりも値が小さくなっていることがわかる。ここから、特に受検者特性値の次元間相関が高い場合の共通受検者に注目した等化については、Hirsch の方法に比較して提案手法が有効な手法となる傾向があるといえそうである。

ただし、表 (5.2), 表 (5.3) を詳しくみると、相関 ρ が大きい場合においても、Hirsch の方法と提案手法で全般に Q1 と Q3 のレンジに重なりが大きいことがわかる。図 (5.1) は、横軸を提案手法で $\lambda = 0.0$ とした場合、縦軸を Hirsch の方法として、 $\rho = 0.0$ の場合の各繰り返しにおける $RMSE(\boldsymbol{\theta})_k$ の値をプロットしたものである。図 (5.1) から、 $RMSE(\boldsymbol{\theta})_k$ については、繰り返し (k) のほぼ全てのケースで Hirsch の方法のほうが提案手法よりも値が小さくなっていることがわかる。図 (5.2), 図 (5.3), 図 (5.4) は同じく $\rho = 0.3, 0.6, 0.9$ とした場合の各繰り返しにおける $RMSE(\boldsymbol{\theta})_k$ の値をプロットしたものである。 $\rho = 0.3$ の場合には提案手法と Hirsch の方法の $RMSE(\boldsymbol{\theta})_k$ がほぼ一致し、 ρ が 0.3 より大きい場合には、各繰り返しについてみても、提案手法のほうが $RMSE(\boldsymbol{\theta})_k$ の値が小さくなることがわかる。

各 ρ の値について、Hirsch の方法と提案手法の等化後の受検者特性値の相関を調べたところ、 ρ の値によらず全ての繰り返しにおいて 0.998 から 0.999 と 1.0 に非常に近い値となっていた。また、実験では、等化先、等化後の受検者特性値の平均は各次元について 0 に中心化されている。したがって、Hirsch の方法と提案手法の $RMSE(\boldsymbol{\theta})$ などの実験結果の違いは、等化後の受検者特性値の次元毎の分散にその要因が存在する可能性が高いものと考えられる。提案手法は等化先と等化後の受検者特性値の分散を一致させるものであるが、相関 ρ が小さい場合には、この制約が却って個々の受検者の特性値の近接性を阻害する方向に機能しているものと考えられる。

さて、図 (5.5) は $\rho = 0.6$ の場合の繰り返し毎の $RMSE(\mathbf{a})_k$ を提案手法と

Hirsch の方法についてプロットしたものである。前述のとおり $\rho = 0.6$ の場合には、Hirsch の方法に比較して提案手法のほうが $RMSE(\theta)_k$ が小さい傾向がみられたが、 $RMSE(\mathbf{a})_k$ では、提案手法のほうが全般には値が小さい傾向が確認できるものの、必ずしも常に提案手法が Hirsch の方法に比較して良好な結果となるわけではないことがわかる。本研究では主に $RMSE(\theta)$ と $RMSE(\mathbf{a})$ に注目して、各手法の平均的な性質について調べるが、ある手法が全ての点において常に他の手法より優れている訳ではないという点は他の ρ の値の場合や、Min の方法についても同様であり、実際の等化の際には各等化法を用いた等化を行い、それぞれの結果を比較して適切な手法を総合的に判断するなど、状況に応じて等化法を使い分けるのが望ましいといえるだろう。

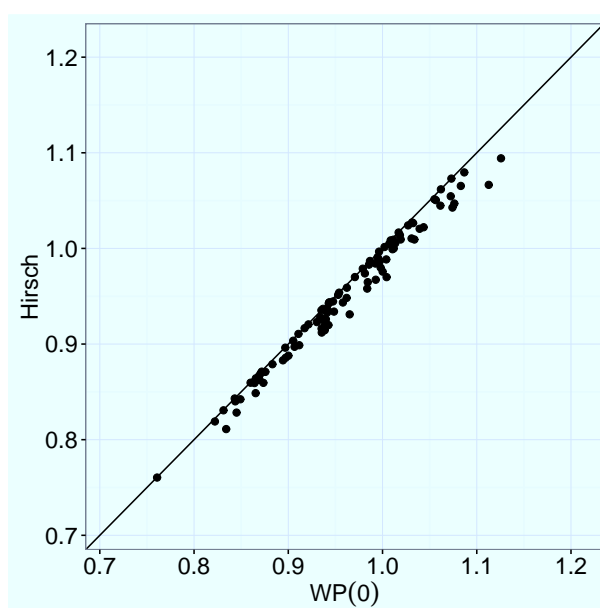


図 5.1: $\rho = .0$ の場合の WP 法 ($\lambda = 0.0$) と Hirsch の方法の $RMSE(\theta)_k$

つぎに、Min の方法と提案手法を比較する。まず表 (5.2) の $RMSE(\theta)$ では相関 ρ 、重み λ の値によらず、Min の方法よりも提案手法のほうが全般に値が小さくなっている。この結果は、Min の方法が共通項目のみに注目した等化法であり受検者特性値の情報を利用しない一方、提案手法は重み $\lambda = 1.0$ とした場合でも共通受検者の情報を利用する方法であることによるものであると考えられる。一方、表 (5.3) の $RMSE(\mathbf{a})$ についてみると、提案手法において $\lambda = 1.0$ として共通項目を重視した場合であっても、Min の方法のほうが全般に値が小さくなっていることがわかる。 $\rho = 0.0$ の場合には、提案手法で $\lambda = 1.0$ のとき $RMSE(\mathbf{a}) = 0.205$ 、Min の方法では 0.190 とその差はそれほど大きくはないが、 $\rho = 0.9$ の場合には、 $RMSE(\mathbf{a})$ は提案手法で 0.368 、Min の方法では 0.179 と Min の方法のほうが大幅に小さい値となっており、

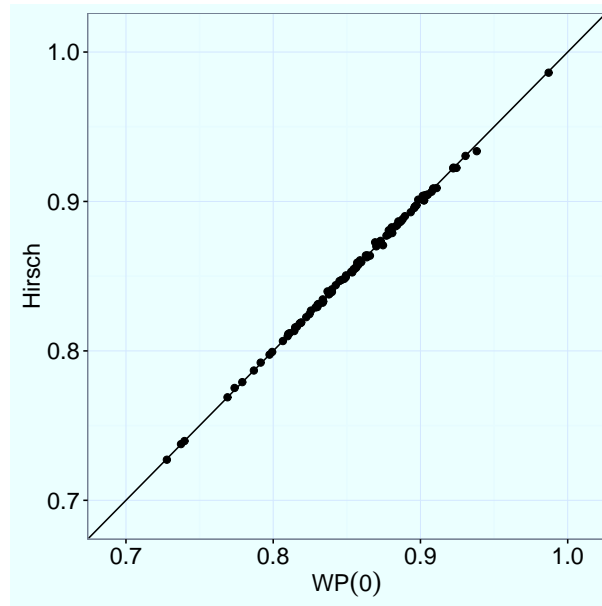


図 5.2: $\rho = .3$ の場合の WP 法 ($\lambda = 0.0$) と Hirsch の方法の $RMSE(\theta)_k$

特に受検者特性値の次元間相関が高く共通項目のパラメータの等化先との近接性が重視される場合には、Min の方法による等化がより有効であるものといえる。ただし $\rho = 0.9$ の場合、Min の方法の $RMSE(\theta)$ は 1.218 と提案手法で $\lambda = 1.0$ とした場合の 0.508 に比較して、かなり大きな値となっている。図 (5.6) は、 $\rho = 0.9$ の場合の Min の方法による 1 回の繰り返しの結果について、横軸を等化後の第 1 次元の、縦軸を等化後の第 2 次元の受検者特性値として、等化後の受検者の散布図を一例として描いたものである。図 (5.7) は同じデータに関する $\lambda = 1.0$ とした場合の提案手法の結果である。これらの図からも Min の方法では受検者特性値の相関や分散が十分に再現できていないことがわかる⁴。これは、Min の方法が基本的に直交回転に基づく方法であり、等化先の受検者特性値の分散や相関に追従することができないことが原因であると思われる⁵。これらの結果から、受検者特性値の相関が大きい場合に共通項目による等化法である Min の方法を利用すると、共通項目の近接性は高いものの受検者特性値の近接性に問題が生じる一方、提案手法は共通項目に加えて受検者特性値の情報を利用することで、両者のバランスをとった等化を実

⁴図 (5.6) および図 (5.7) の等化先の受検者特性値の分散は第 1 次元が 0.92、第 2 次元が 0.90、次元間の相関は 0.90、等化前の分散は第 1 次元が 1.56、第 2 次元が 0.35、相関は 0.26 であった。また、Min による等化後の分散は第 1 次元 1.36、第 2 次元 2.99、相関は 0.62、提案手法による等化後の分散および相関は等化先のものと同じ値であった。

⁵この結果は、等化前のパラメータを promax 法で算出したこと、受検者特性値の事前分布を独立な標準正規分布に設定したことによるものであり、パラメータ算出の際に等化先の受検者特性値の次元間相関を考慮するような手法を利用すれば、別の結果が得られる可能性が高い。ただし、そのことはやはり多次元 IRT の等化において受検者特性値の次元間相関などの情報を考慮することの重要性を示唆するものである。

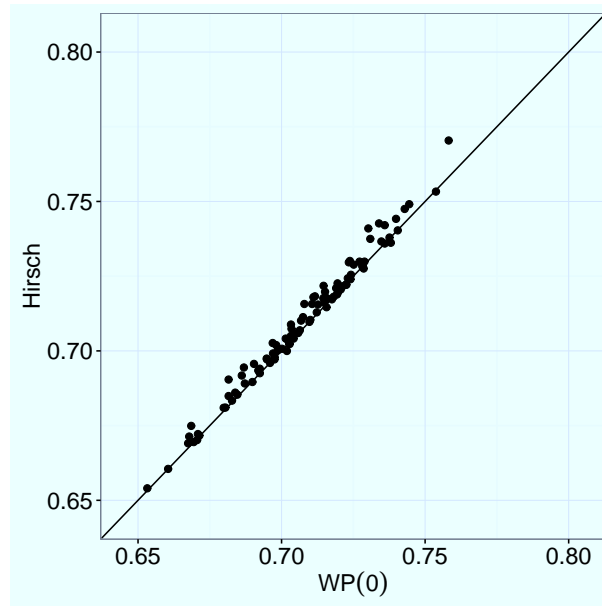


図 5.3: $\rho = .6$ の場合の WP 法 ($\lambda = 0.0$) と Hirsch の方法の $RMSE(\theta)_k$

現するものであるといえる。

5.3.2 次元の数に関する実験

つぎに、次元数の影響を調べるために、次元数 m を $m = 2, 3, 4$ と変化させた場合について実験を行った。その他の設定および手順は相関に関する実験と同様とした。ただし、受検者特性値の次元間の相関は、次元数によらず $\rho = 0.5$ に固定した。たとえば、 $m = 3$ の場合は、第 1 次元と第 2 次元、第 1 次元と第 3 次元、第 2 次元と第 3 次元と、3 つの次元間の相関が考えられるが、これら 3 箇所とも 0.5 とした。提案手法、Hirsch の方法、Min の方法による等化の結果について、手法毎に $RMSE(\theta)$, $RMSE(\mathbf{a})$, $RMSE(S)$ の 3 つの指標の値を確認した。また、 $RMSE(\theta)$ と $RMSE(\mathbf{a})$ については、各手法ごとに次元数 $m = 4$ の場合から $m = 2$ の場合の結果を引いた値を求めた。 $RMSE(\theta)$ の結果を表 (5.5) に、 $RMSE(\mathbf{a})$ の結果を表 (5.6) に、 $RMSE(S)$ の結果を表 (5.7) にそれぞれまとめる。

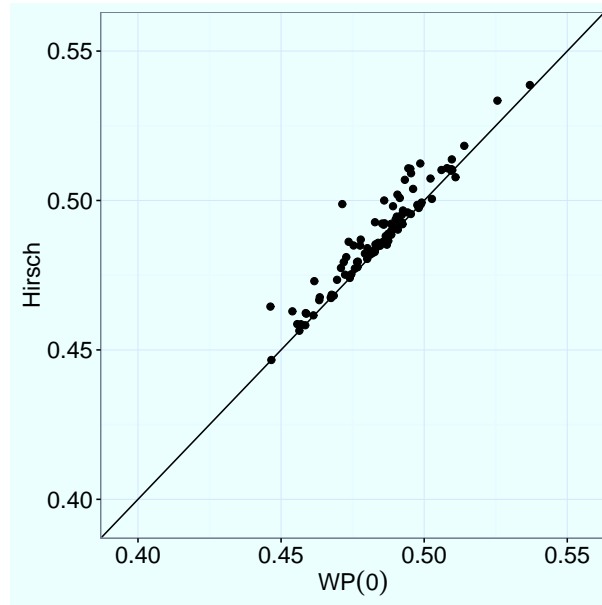


図 5.4: $\rho = .9$ の場合の WP 法 ($\lambda = 0.0$) と Hirsch の方法の $RMSE(\boldsymbol{\theta})_k$

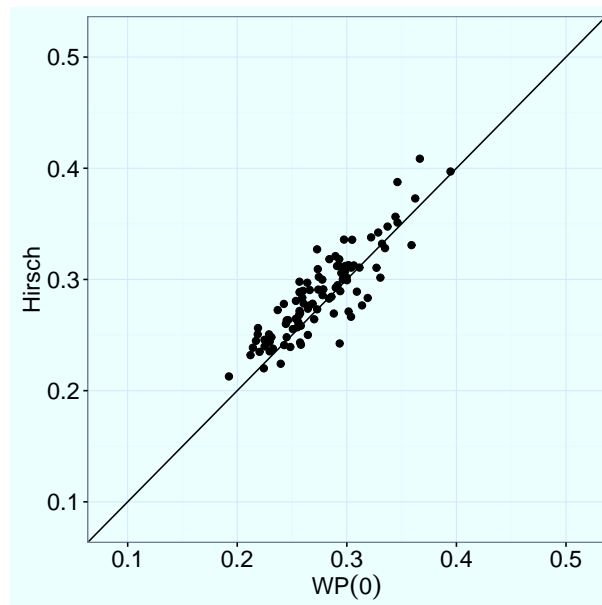


図 5.5: $\rho = .6$ の場合の WP 法 ($\lambda = 0.0$) と Hirsch の方法の $RMSE(\boldsymbol{a})_k$

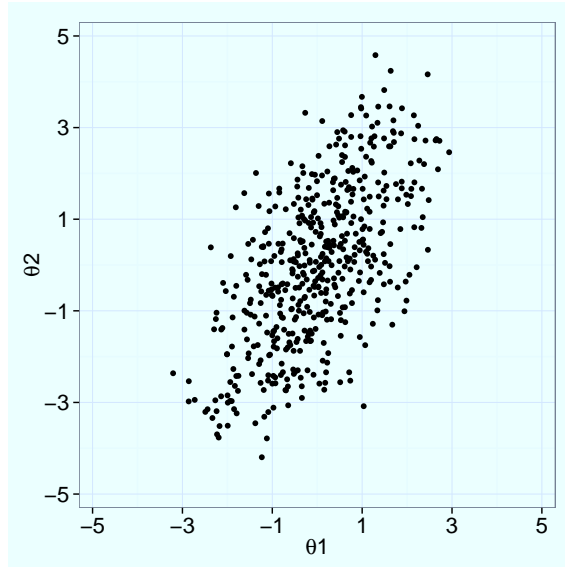


図 5.6: $\rho = .9$ の場合の Min の方法による等化後の受検者特性値の例

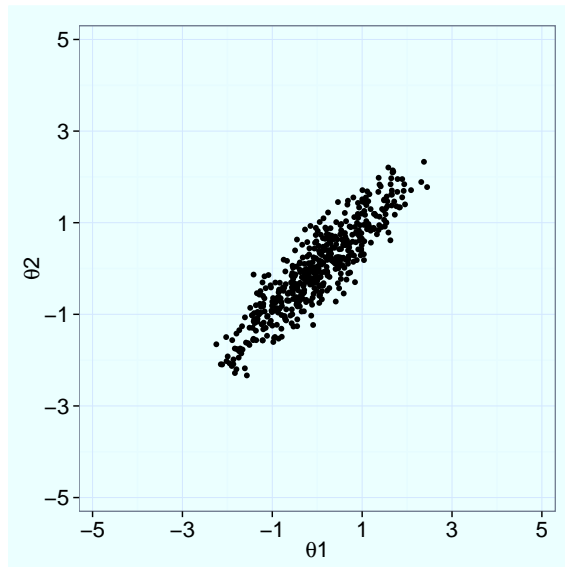


図 5.7: $\rho = .9$ の場合の WP($\lambda = 1.0$) 法による等化後の受検者特性値の例

表 5.5: 次元数 m を変化させた場合の各手法の $RMSE(\theta)$ の値

		WP						
		$\lambda = 1.0$	$\lambda = .75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.0$	Hirsch	
$m = 2$	$RMSE(\theta)$.781	.772	.769	.768	.767	1.169	.768
	Q1	.761	.755	.752	.749	.749	1.029	.750
	中央値	.783	.774	.771	.770	.769	1.135	.770
	Q3	.798	.791	.787	.787	.787	1.262	.786
$m = 3$	$RMSE(\theta)$.832	.818	.813	.810	.809	1.555	.810
	Q1	.810	.800	.795	.792	.791	1.445	.793
	中央値	.825	.817	.813	.809	.808	1.546	.809
	Q3	.841	.831	.827	.825	.825	1.611	.825
$m = 4$	$RMSE(\theta)$.871	.848	.843	.839	.836	1.856	.836
	Q1	.849	.832	.830	.827	.824	1.772	.824
	中央値	.871	.850	.846	.841	.839	1.867	.839
	Q3	.889	.866	.858	.853	.851	1.931	.850
$m = 4$ と $m = 2$ の $RMSE(\theta)$ の差		.090	.076	.074	.071	.069	.687	.068

表 5.6: 次元数 m を変化させた場合の各手法の $RMSE(\mathbf{a})$ の値

		WP						
		$\lambda = 1.0$	$\lambda = .75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.0$	Hirsch	
$m = 2$	$RMSE(\mathbf{a})$.248	.249	.256	.264	.273	.184	.277
	Q1	.229	.230	.237	.243	.247	.170	.257
	中央値	.246	.246	.251	.256	.261	.185	.270
	Q3	.259	.262	.267	.279	.292	.194	.291
$m = 3$	$RMSE(\mathbf{a})$.277	.279	.286	.301	.325	.183	.336
	Q1	.256	.261	.269	.283	.302	.176	.316
	中央値	.276	.277	.282	.293	.315	.183	.330
	Q3	.294	.294	.299	.313	.340	.191	.345
$m = 4$	$RMSE(\mathbf{a})$.311	.315	.321	.340	.389	.181	.393
	Q1	.286	.290	.297	.317	.365	.173	.372
	中央値	.308	.312	.317	.336	.382	.182	.394
	Q3	.329	.328	.331	.349	.404	.186	.408
$m = 4$ と $m = 2$ の	$RMSE(\mathbf{a})$ の差	.063	.066	.065	.076	.117	-.003	.116

表 5.7: 次元数 m を変化させた場合の各手法の $RMSE(S)$ の値

	WP						Min	Hirsch
	$\lambda = 1.0$	$\lambda = .75$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.0$	$\lambda = 0.0$		
$m = 2$.000	.000	.000	.000	.000	.000	1.622	.031
$m = 3$.000	.000	.000	.000	.000	.000	2.632	.034
$m = 4$.000	.000	.000	.000	.000	.000	3.694	.029

まず、表 (5.5) の $RMSE(\theta)$ について提案手法の結果をみると、次元数 m に関わらず λ の値が 0.0 に近いほど $RMSE(\theta)$ の値が小さくなっていることがわかる。また、表 6 では λ が 1.0 に近いほど $RMSE(\mathbf{a})$ の値が小さくなっている。これは、相関に関する実験の場合と同様、重み λ が提案手法において、共通受検者と共通項目に関する等化基準を調整することの効果である。

表 (5.7) の $RMSE(S)$ についてみると、提案手法では m, ρ によらず値が 0 となっている。厳密には、 $RMSE(S)$ は 2 次元の場合 0.4×10^{-16} 、3 次元では 0.5×10^{-16} 、4 次元では 0.6×10^{-16} といずれの場合にも極小さい値であった。提案手法において等化先と等化後の分散と相関係数が一致するという性質は、数値計算に伴う誤差の影響が存在しなければ理論的に常に成り立つものである。計算環境に依存する部分があるが、実験の結果から少なくとも 4 次元程度までであれば実際の計算においても実用上この性質が成り立っていると考えて問題ないものといえるだろう。

さて、あらためて表 (5.5) をみると、いずれの手法でも次元数 m が大きくなるほど、 $RMSE(\theta)$ の値が大きくなる傾向があることがわかる。とくに Min の方法では、この傾向が顕著である。表 (5.6) からは、提案手法と Hirsch の方法について、次元数が大きくなるほど $RMSE(\mathbf{a})$ が大きくなる傾向が確認できる。また、提案手法について、表 (5.5) の次元数 $m = 4$ の場合の $RMSE(\theta)$ から $m = 2$ の場合の $RMSE(\theta)$ を引いた差の値をみると、重み λ が 1.0 に近いほど $m = 2$ の場合と $m = 4$ の場合の $RMSE(\theta)$ の乖離が大きくなっていることがわかる。表 (5.6) では、重み λ が 0.0 に近いほど $m = 2$ の場合と $m = 4$ の場合の $RMSE(\mathbf{a})$ の差が大きくなっている。

これらの結果から、次元数大きい場合、共通項目に注目した方法では共通受検者の、共通受検者に注目した方法ではもう一方の共通項目の等化後の値が等化先と乖離しやすい傾向があることが読み取れる。したがって、次元数大きい場合には特に、共通項目、共通受検者、双方の等化基準について注意深く確認する必要があるものといえる。また、提案手法は、重み λ がいずれの値の場合でも、 $RMSE(\theta)$ 、 $RMSE(\mathbf{a})$ の双方とも、概ね結果が Min の方法と Hirsch の方法の中間的な値となっており、次元数が増えた場合でも、総合的には共通項目、共通受検者の両方の基準とも極端には悪化しにくい傾向があることがわかる。この結果は、Min の方法、Hirsch の方法に比較して、提案手法によって相対的に安定性の高い等化が実現できることを表すものと解釈できる。

5.4 まとめと課題

本章では、多次元 IRT において共通項目による等化と共通受検者による等化の両方が実行可能な場合に対応した、項目パラメータに関する基準と受検者特性値による基準を統合した新たな等化法を提案した。この際、受検者特性値の次元間の相関係数を等化先のテストと等化後のテストで一致させること

を制約として考え、常にこれを満たすような方法を構成し、計算機を利用してその特性を既存手法と比較した。提案手法は、重み λ によって共通項目による基準と受検者特性値による基準を統一的に取り扱うことを目指すものであり、 λ はどちらの基準を重視するかを調整するような機能をもつと考えられる。実験の結果、重み λ によって、概ねこのような調整が実現できることが確認できた。また、提案手法において $\lambda = 0.0$ として受検者特性値だけに注目した場合、特に受検者特性値の次元間の相関が大きいときに、共通受検者による既存の等化法である Hirsch の方法よりも、受検者特性値、項目パラメータの等化先との近接性という面で良好な結果が得られることが把握できた。加えて、提案手法は重み λ の値によらず、共通項目による等化法である Min の方法に比較して、受検者特性値について、より近接性の高い結果をもたらすことが確認できた。受検者特性値の次元が大きい場合も含めて、提案手法は項目に関する情報と受検者に関する情報の両方を利用することで、両者をバランスよく考慮した等化を実現するものであるといえる。

提案手法は前章で示した共通受検者デザインの場合の重み付きプロクラステス回転法を共通項目の情報も利用する方向に拡張し、共通受検者による等化基準と共通項目による等化基準を重み λ を用いて統合したものとみなすことができる。基準間の重み $\lambda = 1$ とした場合には共通項目に関するウェイトが最大となるが、実験でも確認されたように提案手法ではこの場合にも等化先と等化後の受検者特性値の分散共分散行列が一致する。つまり、共通項目に注目した等化の場合でも共通受検者の分散共分散行列という形でその情報を利用する点に特徴がある。このことは、提案手法を利用するためには項目パラメータに加えて、受検者特性値に関する情報が必要となることを意味するが、特に $\lambda = 1$ の場合には等化先と等化対象のテストの受検者特性値の分散共分散行列が存在すれば個々の受検者の特性値は存在しなくても共通項目を利用した等化が実現できる。もちろん、個々の受検者の特性値が存在する場合には、どのような重みの場合にも提案手法による等化が実現できる。

さて、等化先のテストから直接計算した相関係数は一般に一定の誤差を含んでおり真値そのものではない。等化後の受検者特性値について、等化先のテストの推定値の相関ではなく、受検者特性値の真の相関構造を再現することを重視するような立場も考えられる。ところで、提案手法では分散共分散行列にターゲットを指定するが、ターゲットとなる分散共分散行列は、必ずしも等化先のテストのデータから計算したものである必要はなく、たとえば過去の研究の蓄積などによって次元間の相関が明確に把握できている場合には、これを利用することが可能である。特に受検者数が少ない場合や推定に関連する誤差が大きい場合など等化先のテストの相関係数が不安定であることが懸念される際には、このような方法を検討する価値があると思われる。

第6章 実データを用いた提案手法 の評価

この章では、OECD 学習到達度調査 PISA の実際の回答データに対して受検者特性値に注目した重み付きプロクラステス回転法による等化を適用し、その結果を確認する。PISA は、15 歳児を対象として 3 年ごとに多数の国と地域にわたって実施される大規模な学力調査であり、国際数学・理科教育動向調査 TIMSS などと並んでその結果は本邦でも広く注目されている。ここでは、2012 年の日本のデータのうち、数学的リテラシーと科学的リテラシーの項目に対する回答データを利用して検討を行う¹。なお PISA では、ブックと呼ばれる問題項目の構成の異なるいくつかのテスト冊子が用意されている。ここでは、第 1 ブックの受検者 482 人を分析の対象とした。

6.1 分析手順

分析の手順はつぎのとおりである。まず、数学的リテラシー 25 問と科学的リテラシー 38 問について、予備的に 1 次元の 2 パラメータ・ロジスティックモデル、2 次元の 2 パラメータ・ロジスティック補償型モデルを適用して項目パラメータを算出し、相対的に項目数の多い科学的リテラシーを中心に、識別力パラメータの小さい項目や、識別力が異常に大きく推定される項目を除外した。分析対象とした項目の一覧を表 (6.1) にまとめる。

つぎに、共通受検者による等化を実現するために、問題項目を 2 群に振り分け仮想的な 2 つのテストフォームを構成した。以降これらを form1, form2 と呼ぶこととする。form1, form2 とともに数学的リテラシーと科学的リテラシーの項目を含み、問題数は form1 が 35 問、form2 は 34 問である。なお、後に項目パラメータについて等化精度を検討するために、数学的リテラシー 10 問、科学的リテラシー 10 問の計 20 問が form1 と form2 に共通して含まれるようにした²。各フォームに含まれる問題項目を表 (6.2) および表 (6.3) にまとめた。

いま、2 つのテストフォームは、問題項目は異なるが同一の潜在特性、数学的リテラシー、科学的リテラシーを測定するものであり、また共通する受検者を有していることから、共通受検者による等化が可能な状態である。なお、多次元 IRT ではそれぞれのテストフォームに含まれる問題項目が、その程度は別にして数学的リテラシーと科学的リテラシーの両方に関連している可能性があるものとして分析を行う。したがって、各項目が元々どちらの領域を想定して作成されたかということに関わらず、手続き的には各フォームに対する問題項目をひとつのまとまりとして分析を行う³。

¹最新の PISA は 2015 年のものであるが、本章執筆時点でこのデータは公開されていなかったため、入手可能なものうち最も新しい 2012 年のデータを利用した。

²本章では共通受検者による等化について検討しており、テストフォームに共通して含まれる項目は等化には利用せず、専ら等化精度の検討のために利用する点を念のため確認しておく。

³ただし、各問題項目がどのような構成概念を想定して作成されたかということは、非常に重要な情報である。多次元 IRT の等化においてこの情報を考慮する方法の可能性については、第 7 章で触れる。

形式的に 482 人全員を共通受検者と捉え, form1 をターゲットとなる等化先のテストフォーム, form2 を等化対象のテストフォームとして等化を行う. このためにまず, form1, form2 のそれぞれについて 2 パラメータ・ロジスティック補償型モデルを適用して受検者特性値と項目パラメータを算出した. この際, form1 については promax 解を, form2 については varimax 解を採用し, 次元数はそれぞれ 2 とした⁴.

form1 について, 横軸を第 1 次元, 縦軸を第 2 次元の受検者特性値として受検者をプロットしたものが図 (6.1), form2 について同様に受検者をプロットしたものが図 (6.2) である. 2 つの図は同じ受検者群の特性値を示すものである. 同じ受検者については, form1, form2 のどちらを利用した場合であっても, スコアが同一となることが期待されるが, これらの図から明らかなように, 単純に受検者特性値を推定した場合にはパラメータ推定の不定性などの影響で, 通常そのようにはならない⁵. したがって, 2 つのテストフォームのスコアが近接するように等化を行うことが必要となる. 等化法としては第 4 章で示した受検者特性値に注目した重み付きプロクラステス回転法を利用する. この方法は第 5 章の共通受検者と共通項目の両方を利用する重み付きプロクラステス回転による等化法において, 重み λ の値を 0.0 とした場合に相当する. なお, この手法では等化先のテストフォームと等化後のテストフォームの相関構造が一致する特性があり, form1 によって推定された受検者特性値の次元間の相関と, form2 を等化した後の次元間の相関が一致する. 比較のために共通受検者による等化法である Hirsch の方法を利用した場合についても確認した.

6.2 結果

form2 について Hirsch の方法によって等化を行った場合の等化後の受検者特性値について, 横軸を第 1 次元, 縦軸を第 2 次元として受検者をプロットしたものが図 (6.3), 同じく重み付きプロクラステス法 (WP) の結果が図 (6.4) である. 図 (6.5), 図 (6.6) はターゲットである form1 の受検者特性値を横軸, Hirsch の方法による等化結果を縦軸として, それぞれ第 1 次元, 第 2 次元の結果をプロットしたものである. 同様に重み付きプロクラステス回転の結果を図 (6.7), 図 (6.8) に示した. また, それぞれの等化法についてつぎの $RMSE(\theta)$ を計算した. 結果をまとめたものが表 (6.6) である. また, 参考のために form1 の受検者特性値と, 等化後の対応する次元の受検者特性値を表 (6.7) にまと

⁴2 つのフォームで多くの項目を共通としたため, フォーム毎の受検者特性値の差が明確になるよう form1 と form2 で異なる解を採用した.

⁵form1 の受検者特性値の次元間の相関係数は, .9411, form2 の等化前の受検者特性値の次元間の相関係数は, -.5916 であった. なお, それぞれ 1 次元の IRT で分析した場合の相関は 0.79 であった.

めた.

$$RMSE(\boldsymbol{\theta}) = \sqrt{\frac{1}{mn} \sum_{i=1}^n \sum_{l=1}^m (\hat{\theta}_{gil} - \theta_{il})^2} \quad (6.1)$$

上式で, m は次元数を, n は受検者数を表す. また, $\hat{\theta}_g$ は等化後の form2 の受検者特性値を, θ はターゲットである form1 の受検者特性値を表す.

PISA のデータは次元間の相関が高いこともあり, 散布図からは等化法ごとの傾向を明確に把握するのは難しいが, $RMSE(\boldsymbol{\theta})$ をみると, Hirsch の方法よりも重み付きプロクラステス回転法による等化結果のほうが等化後の受検者特性値について, ターゲットである form1 に対する近接性が若干高い結果となっていることがわかる. 第 4 章, 第 5 章の計算機実験の結果から, 等化先の受検者特性値の次元間の相関が高い場合には特に, 重み付きプロクラステス回転法の受検者特性値の等化先に対する追従性は, Hirsch の方法に比較して高くなる傾向があることが示された. 上記の結果はこれに合致するものである.

なお, 等化係数行列 W は Hirsch の方法を利用した場合,

$$W_{Hirsch} = \begin{bmatrix} -0.934 & -0.481 \\ 0.382 & 0.850 \end{bmatrix} \quad (6.2)$$

重み付きプロクラステス回転の場合は,

$$W_{WP} = \begin{bmatrix} -0.918 & -0.453 \\ 0.414 & 0.868 \end{bmatrix} \quad (6.3)$$

であった.

等化後の受検者特性値の次元間相関についてみると, 重み付きプロクラステス回転法による form2 の等化後の受検者特性値の次元間の相関は .9411 であった. これは form1 のそれと同じ値である. 重み付きプロクラステス回転法による等化では, 等化先と等化後の相関係数が必ず一致する性質がある. この結果は重み付きプロクラステス回転法の特徴を示すものである. Hirsch の方法の場合には, 等化後の次元間の相関は .9410 であった. Hirsch の方法では, 等化後の次元間相関は等化先のものとは必ずしも一致はしないが, ここで取り上げたデータについては, 等化先と非常に近い値となっている. 第 4 章における, 次元間の相関係数に関する実験でも, 等化先のテストの次元間相関が高い場合には Hirsch の方法の相関構造の追従性が高くなることが示されており, この結果はシミュレーション実験の結果とやはり合致するものである.

つぎに, 項目パラメータの等化結果について確認する. 表 (6.8) は, form1 と form2 の両方に共通して含まれる 20 の項目について, 等化法毎の等化後の \boldsymbol{a} パラメータの値をまとめたものである.

等化後の \boldsymbol{a} パラメータの値について, 等化先である form1 の値とのギャップを表すつぎの $RMSE(\boldsymbol{a})$ の値をまとめたのが, 表 (6.9) である.

$$RMSE(\mathbf{a}) = \sqrt{\frac{1}{mp} \sum_{j=1}^p \sum_{l=1}^m (\hat{\theta}_{gjl} - \theta_{jl})^2} \quad (6.4)$$

上式で, p は 2 つのフォームに共通して含まれる項目数を表す. 他の記号は $RMSE(\boldsymbol{\theta})$ と同様である.

$RMSE(\mathbf{a})$ についても, 受検者特性値に注目した重み付きプロクラステス回転法のほうが値が小さくなっており, Hirsch の方法に比較して等化後の項目パラメータがターゲットに近い値となっていることがわかる. これも, 第 4 章, 第 5 章のシミュレーション実験に一致する結果である.

念のため, form1 の同じ項目に対して各等化法による等化後の form2 の項目特性値がどの程度近接しているかを調べるために, 項目別につぎの $RMSE(\mathbf{a})_j$ を求めた.

$$RMSE(\mathbf{a})_j = \sqrt{\frac{1}{m} \sum_{l=1}^m (\hat{\theta}_{gjl} - \theta_{jl})^2} \quad (6.5)$$

結果をまとめたものが, 図 (6.9) である.

表 6.1: PISA2012 日本の第 1 冊子の問題項目と本研究の分析対象項目

連番	問題項目	領域	分析対象
1	PM00F Q1	数学	*
2	PM00G Q1-Q3	数学	*
3	PM909 Q1-Q3	数学	*
4	PM918 Q1,Q2,Q5	数学	*
5	PM923 Q1,Q3,Q4	数学	*
6	PM924 Q2	数学	*
7	PM949 Q1-Q3	数学	*
8	PM955 Q1-Q3	数学	*
9	PM995 Q1-Q3	数学	*
10	PM998 Q2,Q4	数学	*
11	PS131 Q2,Q4	科学	-
12	PS256 Q1	科学	*
13	PS326 Q1-Q4	科学	*
14	PS413 Q4-Q6	科学	*
15	PS415 Q2,Q7,Q8	科学	*
16	PS425 Q2-Q5	科学	*
17	PS428 Q1,Q3,Q5	科学	*
18	PS438 Q1-Q3	科学	-
19	PS465 Q1,Q2,Q4	科学	-
20	PS478 Q1-Q3	科学	*
21	PS498 Q2-Q4	科学	-
22	PS514 Q2-Q4	科学	*
23	PS131 Q2-Q4	科学	-
24	PS438 Q3	科学	*

表 6.2: テストフォームへの項目の割付 (数学的リテラシー)

連番	項目 ID	Form1	Form2
1	PM00FQ01	*	*
2	PM00GQ01	*	*
3	PM903Q01	*	*
4	PM903Q03	*	*
5	PM909Q01	*	*
6	PM909Q02	*	*
7	PM909Q03	*	*
8	PM918Q01	*	*
9	PM918Q02	*	*
10	PM918Q05	*	*
11	PM923Q01	*	
12	PM923Q03		*
13	PM923Q04	*	
14	PM924Q02		*
15	PM949Q01	*	
16	PM949Q02		*
17	PM949Q03	*	
18	PM955Q01		*
19	PM955Q02	*	
20	PM955Q03		*
21	PM995Q01	*	
22	PM995Q02		*
23	PM995Q03	*	
24	PM998Q02		*
25	PM998Q04	*	

表 6.3: テストフォームへの項目の割付 (科学的リテラシー)

連番	項目 ID	Form1	Form2
26	PS256Q01	*	*
27	PS326Q01	*	*
28	PS326Q02	*	*
29	PS326Q03	*	*
30	PS326Q04	*	*
31	PS413Q04	*	*
32	PS413Q05	*	*
33	PS413Q06	*	*
34	PS415Q02	*	*
35	PS415Q07	*	*
36	PS415Q08	*	
37	PS425Q02		*
38	PS425Q03	*	
39	PS425Q04		*
40	PS425Q05	*	
41	PS428Q01		*
42	PS428Q03	*	
43	PS428Q05		*
44	PS478Q01	*	
45	PS478Q02		*
46	PS478Q03	*	
47	PS514Q02		*
48	PS514Q03	*	
49	PS514Q04		*

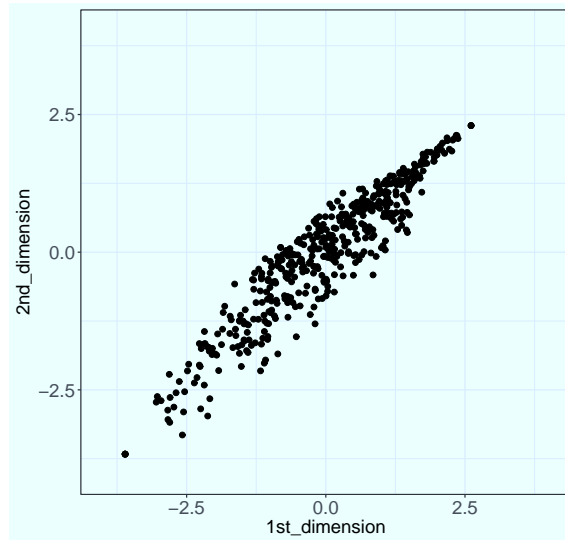


図 6.1: form1 の受検者特性値の散布図 (横軸:第 1 次元, 縦軸:第 2 次元)

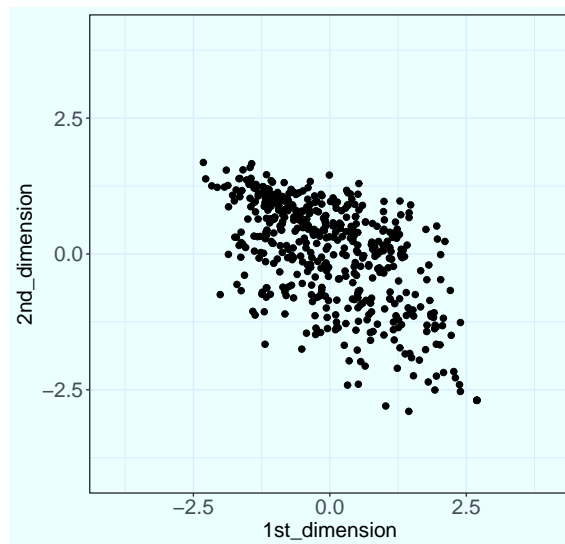


図 6.2: form2 の等化前の受検者特性値の散布図 (横軸:第 1 次元, 縦軸:第 2 次元)

表 6.4: Form1 の項目特性値 (\mathbf{a} パラメータ)

項目 ID	a1	a2
PM00FQ01	1.218	.557
PM00GQ01	1.957	-.369
PM903Q01	2.057	.241
PM903Q03	1.588	.124
PM909Q01	.682	.802
PM909Q02	2.200	-.678
PM909Q03	3.902	-.838
PM918Q01	.409	.772
PM918Q02	.513	.827
PM918Q05	.416	1.036
PM923Q01	1.495	-.049
PM923Q04	2.069	-.198
PM949Q01	2.099	-.340
PM949Q03	1.153	.022
PM955Q02	1.267	-.059
PM995Q01	1.647	.274
PM995Q03	1.389	-.015
PM998Q04	.730	-.023
PS256Q01	-.685	2.41
PS326Q01	-1.165	3.696
PS326Q02	-.155	3.771
PS326Q03	1.177	.907
PS326Q04	.659	.148
PS413Q04	.622	1.003
PS413Q05	.403	1.275
PS413Q06	1.082	.642
PS415Q02	1.013	.829
PS415Q07	-.158	1.049
PS415Q08	.948	.945
PS425Q03	.803	.590
PS425Q05	-.177	1.246
PS428Q03	1.133	.627
PS478Q01	.147	.244
PS478Q03	.202	1.123
PS514Q03	1.306	.128

表 6.5: Form2 の等化前の項目特性値 (\mathbf{a} パラメータ)

項目 ID	a1	a2
PM00FQ01	-1.454	1.036
PM00GQ01	-1.738	.525
PM903Q01	-1.846	1.145
PM903Q03	-1.470	.851
PM909Q01	-1.011	1.046
PM909Q02	-1.749	.354
PM909Q03	-2.578	.933
PM918Q01	-.777	.678
PM918Q02	-.836	.915
PM918Q05	-1.021	.941
PM923Q03	-.689	.500
PM924Q02	-2.223	.932
PM949Q02	-1.466	.524
PM955Q01	-.571	.233
PM955Q03	-1.581	.953
PM995Q02	-1.793	.957
PM998Q02	-.839	.607
PS256Q01	-.454	1.849
PS326Q01	-.587	2.791
PS326Q02	-1.59	3.462
PS326Q03	-1.476	1.468
PS326Q04	-.611	.536
PS413Q04	-.984	1.132
PS413Q05	-.849	1.368
PS413Q06	-1.292	1.051
PS415Q02	-1.315	1.095
PS415Q07	-.375	.823
PS425Q02	-.923	1.318
PS425Q04	-.542	.902
PS428Q01	-1.374	.736
PS428Q05	-1.07	1.285
PS478Q02	-.985	1.324
PS514Q02	-1.475	1.746
PS514Q04	-1.54	1.292

表 6.6: $RMSE(\theta)$

	Hirsch 法	WP 法
全体	.311	.309
第 1 次元	.347	.343
第 2 次元	.271	.271

表 6.7: form1 と等化後の form2 の (θ) の相関

	Hirsch 法	WP 法
第 1 次元	.964	.965
第 2 次元	.975	.975

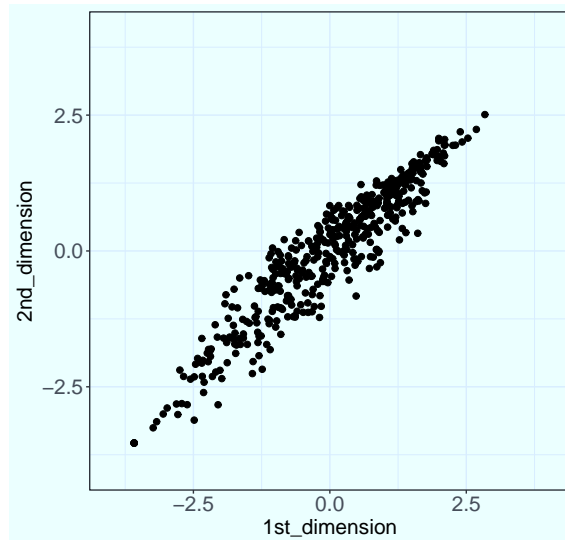


図 6.3: Hirsch の方法による等化後の form2 の受検者特性値の散布図 (横軸: 第 1 次元, 縦軸: 第 2 次元)

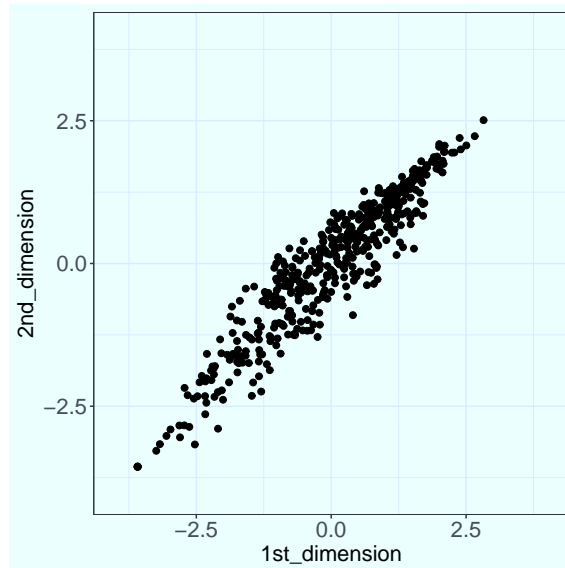


図 6.4: WP 法による等化後の form2 の受検者特性値の散布図 (横軸:第 1 次元, 縦軸:第 2 次元)

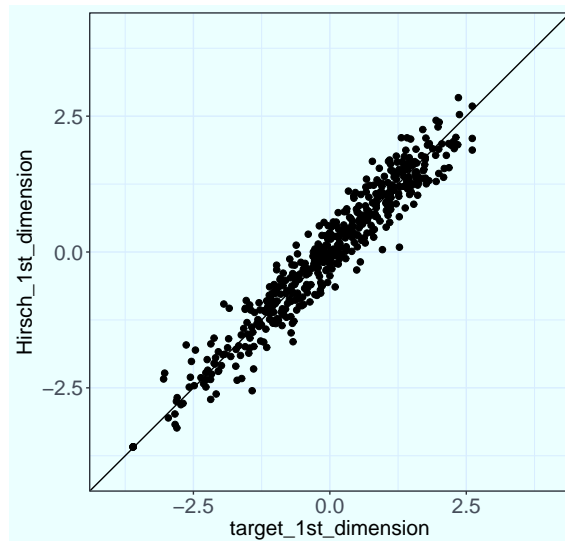


図 6.5: Hirsch の方法による等化後の form2 と form1 の受検者特性値の散布図 第 1 次元 (横軸:form1 の受検者特性値, 縦軸:等化後の form2 の受検者特性値)

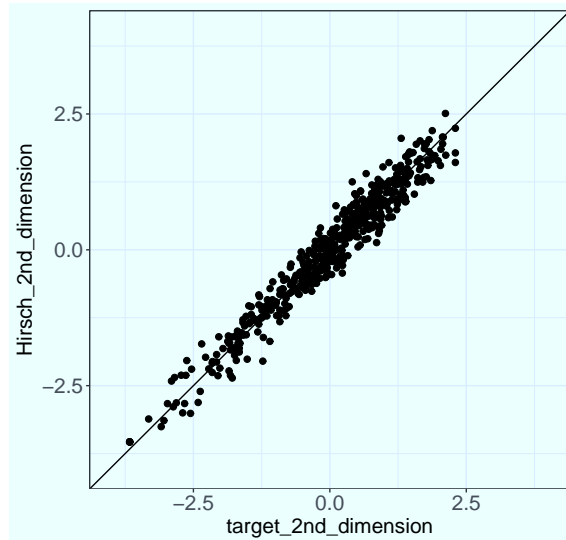


図 6.6: Hirsch の方法による等化後の form2 と form1 の受検者特性値の散布図 第 2 次元 (横軸:form1 の受検者特性値, 縦軸:等化後の form2 の受検者特性値)

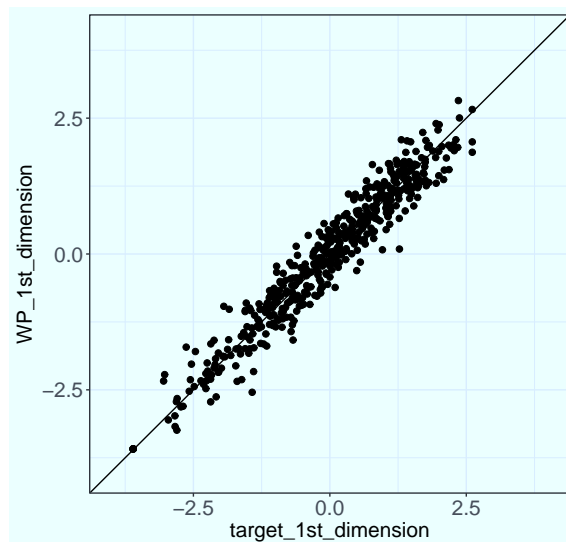


図 6.7: WP 法による等化後の form2 と form1 の受検者特性値の散布図 第 1 次元 (横軸:form1 の受検者特性値, 縦軸:等化後の form2 の受検者特性値)

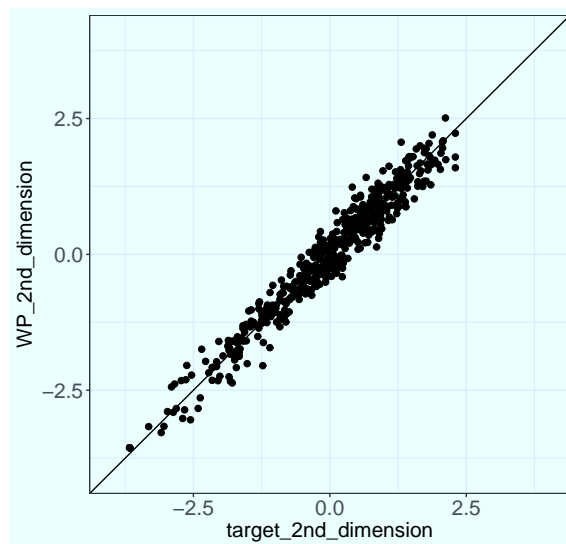


図 6.8: WP 法による等化後の form2 と form1 の受検者特性値の散布図 第 2 次元 (横軸:form1 の受検者特性値, 縦軸:等化後の form2 の受検者特性値)

表 6.8: Hirsh 法と WP 法による Form2 の等化後の項目特性値 (\mathbf{a} パラメータ)

項目 ID	Hirsch 法		WP 法	
	a1	a2	a1	a2
PM00FQ01	1.210	.676	1.302	.572
PM00GQ01	2.008	-.283	2.087	-.390
PM903Q01	1.67	.598	1.78	.470
PM903Q03	1.378	.383	1.463	.283
PM909Q01	.584	.968	.663	.888
PM909Q02	2.158	-.552	2.23	-.656
PM909Q03	2.857	-.184	2.981	-.347
PM918Q01	.548	.551	.603	.493
PM918Q02	.444	.877	.511	.810
PM918Q05	.681	.801	.756	.723
PM923Q03	.566	.334	.610	.285
PM924Q02	2.363	.036	2.475	-.107
PM949Q02	1.63	-.115	1.700	-.207
PM955Q01	.612	-.001	.641	-.037
PM955Q03	1.452	.469	1.545	.361
PM995Q02	1.744	.343	1.844	.223
PM998Q02	.691	.404	.745	.344
PS256Q01	-.824	2.545	-.727	2.477
PS326Q01	-1.381	3.903	-1.238	3.805
PS326Q02	-.512	4.302	-.307	4.134
PS326Q03	.900	1.323	1.012	1.208
PS326Q04	.429	.438	.472	.392
PS413Q04	.479	1.117	.561	1.036
PS413Q05	.105	1.562	.193	1.484
PS413Q06	.972	.800	1.06	.705
PS415Q02	.969	.853	1.06	.756
PS415Q07	-.126	1.025	-.077	.985
PS425Q02	.248	1.439	.336	1.358
PS425Q04	.044	1.041	.102	.990
PS428Q01	1.334	.267	1.411	.175
PS428Q05	.478	1.297	.570	1.208
PS478Q02	.329	1.41	.420	1.325
PS514Q02	.679	1.749	.804	1.627
PS514Q04	1.128	1.014	1.234	.899

表 6.9: $RMSE(\mathbf{a})$

	Hirsch 法	WP 法
全体	.290	.231
第 1 次元	.304	.252
第 2 次元	.275	.207

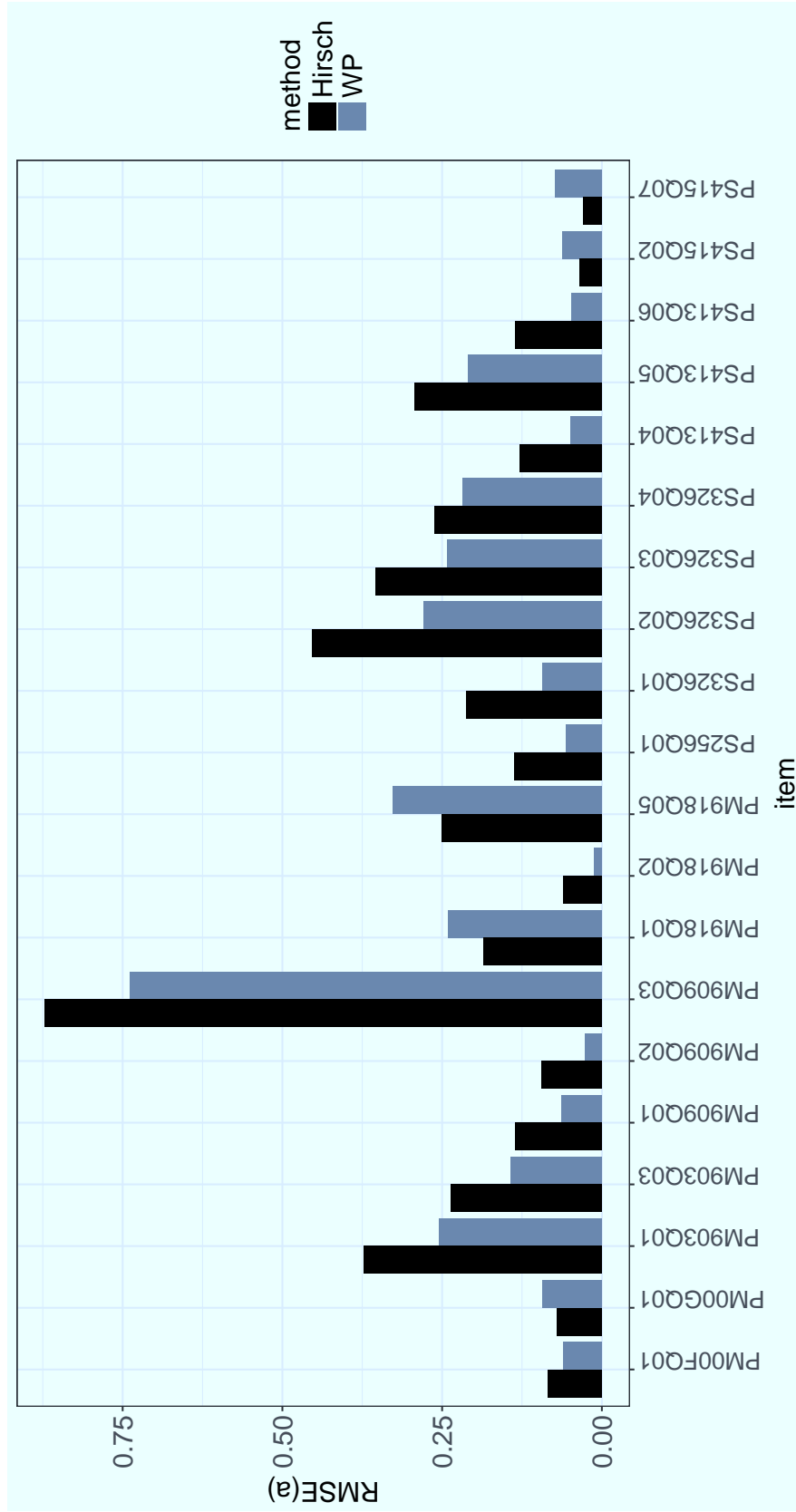


図 6.9: 項目ごとの $RMSE(\alpha)_j$

図から、多くの項目については重み付きプロクラステス回転法による場合のほうが、Hirschの方法に比して $RMSE(\mathbf{a})_j$ の値が小さくなっており、一般に form1 との近接性が高いことがわかる。ただし、Hirschの方法のほうが $RMSE(\mathbf{a})_j$ が小さい項目もいくつか存在し、項目単位でみた場合には、一貫していずれかの方法が有効であるというわけではないことが把握できる。

どのような項目において、Hirschの方法と重み付きプロクラステス回転法で等化結果に違いが出やすいのか調べるために、form1 と等化前の form2 それぞれについて、横軸を第1次元の、縦軸を第2次元の \mathbf{a} パラメータの値として、 $RMSE(\mathbf{a})_j$ が Hirschの方法のほうが小さい項目と重み付きプロクラステス回転法のほうが小さい項目で記号を変えて項目をプロットしたものが、つぎの図(6.10)と図(6.11)である。

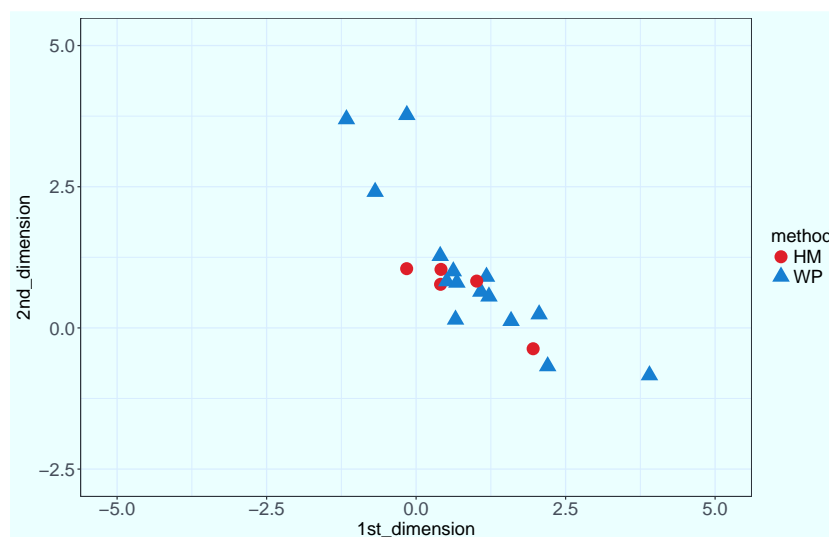


図 6.10: form1 の \mathbf{a} パラメータによる項目のプロット

図から、Hirschの方法で $RMSE(\mathbf{a})_j$ が小さくなる項目に完全に一貫した特徴のようなものは読み取れないが、原点付近にそのような項目が集まっていることがわかる。本章で取り上げた Hirschの方法、受検者特性値に注目した重み付きプロクラステス回転法は共に、項目特性値については明示的には考慮しない方法であり、各回転法の特徴が項目特性値の等化結果に及ぼす影響が理論的に明確なわけではない。上記の傾向が一般的な特徴を表すのかどうかは、単一のデータセットを利用した本章の研究から正確に把握することは難しいが、各等化法の特徴をあらわす仮説として今後検証していく価値があるものと思われる。

さて、本章では、PISA の実際のデータを用いて、提案手法による等化を行った結果を確認した。Hirschの方法との比較において、シミュレーション実験の

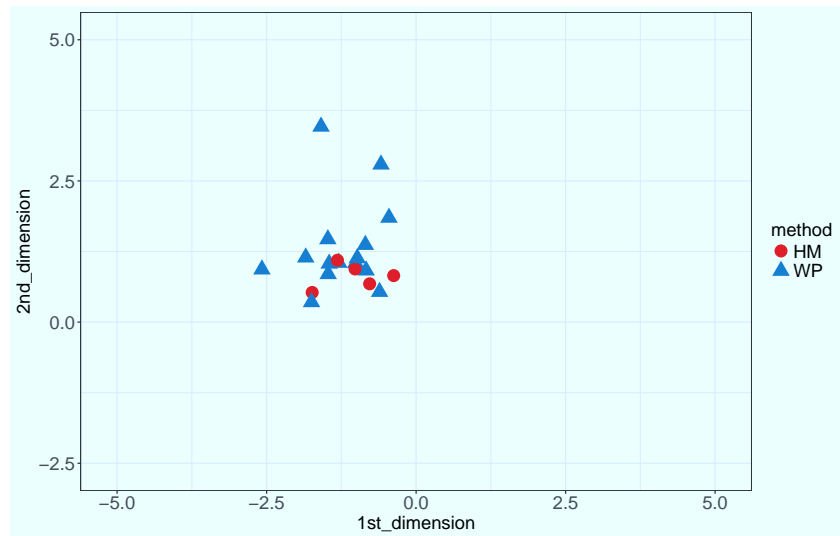


図 6.11: form2 の a パラメータによる項目のプロット

場合と同様, 受検者特性値, 項目特性値の双方とも, よりターゲットである等化先との近接性が高くなる傾向を確認することができた. 本章の結果は, 第 4 章, 第 5 章で示したシミュレーションの結果とも一致するものであり, シミュレーション研究の妥当性を示すものとして理解できる.

ただし本章の分析では, 項目特性値の等化結果の傾向を確認するために, 等化先と等化対象のテストフォームについて約半数の項目を共有する設計とした. この条件は, 等化手法間では共通であるため, 分析の結論に対する影響は大きくないと考えているが, 今回の分析において, 等化後の受検者特性値が通常の共通受検者による等化の場合よりも, 等化先に近接しやすいものとなっている可能性は高い.

また, ここで取り扱ったデータは PISA のデータの一部を取り出したもので, 1 種類のみであり, 実際のデータを利用した場合の提案手法の一般的な傾向を把握するためには, より多くの実際のテストについて同種の検討を行う必要があることは言を待たない. 特に, 受検者特性値の等化結果について, 先行研究と提案手法の間の差異は小さく, この結果をもって一般の場合の手法間の優劣を判断することは難しいものと考えられる. 結果の一般化に向けて, 今後様々なデータについて分析を進めていく必要がある.

第7章 全体のまとめ

7.1 本研究の目的

多次元 IRT の共通受検者による等化法の提案

テストを採点・分析するための理論的な枠組みである IRT のモデルのうち、複数の能力特性が同時に回答に影響を及ぼすことを想定したモデルとして多次元 IRT がある。多くのテストは多次元性を内在させていると考えられる (Ackerman *et al.*, 2005; Yao & Boughton, 2009) ため、多次元 IRT はテストの分析において有望なモデルのひとつであると考えられる。

さて、複数のテストフォームのスコアを、同一スケール上で比較可能とするための変換操作である等化は、多次元 IRT によるテストの実用において重要な課題である。その一方で、多次元 IRT の等化に関する研究は比較的新しい研究分野であり (Simon, 2008)、等化法に関する研究が十分でないことが多次元 IRT の実用化の障害になっているとの意見がある (Min, 2007)。多次元 IRT の主な等化のデザインには、共通項目によるものと共通受検者によるものの 2 つがあるが、このうち共通項目による等化は、複数のテストフォームに共通して含まれる項目について、その特性がテストフォーム間で変化しないことを仮定し、これをアンカーとしてテストフォームのスケール等を調整する操作である。もう一方の共通受検者による等化は、複数のテストフォームに共通する受検者の特性が、受検時点で変化しないことを仮定してスケール等を調整する操作を指す。共通項目による等化と共通受検者による等化は、1 次元の IRT の運用においても共に広く利用される方法であるが、多次元 IRT で等化を行う場合には次元の回転等の問題を考慮する必要があるため、多次元 IRT に独自の等化法が必要となる。

2 つの等化デザインに対応した多次元 IRT の等化法のうち、共通項目による等化については、特徴の把握が十分でない面があるものの、いくつかの等化法が既に提案されている。一方の共通受検者による等化では、先行研究として Hirsch によるもの (Hirsch, 1989) が存在するが、他にはほとんど見当たらないのが現状である。したがって、特に共通受検者デザインによる等化法の研究が、多次元 IRT の応用上重要な意味をもつと考えられる。

さて、共通受検者デザインにおける既存の等化法である Hirsch の方法については、つぎの 2 点を課題として指摘できる。まず、Hirsch の方法は基本的に直交回転に基づく方法であるが、直交性は等化において一種の制約として機能するため、共通受検者について等化先の受検者特性値と等化後の受検者特性値をできるだけ近接させることが重視される場合に、この方法では両者の近接性が十分なものとならない可能性が高い点である。2 点目として、Hirsch の方法による等化後の共通受検者の特性値の相関構造は、等化先のそれとは完全には一致しないが、次元間の相関構造は当該テストが何を測定しているかということに関連する重要な指標であり、この構造が等化先と一致しないことはテストの運用上の不都合につながる可能性がある点である。

本研究では、これらの課題を考慮した上で、多次元 IRT における共通受検者デザインによる新たな等化法を提案し、その特性を計算機実験および実データを用いた分析によって把握することを目指した。

共通項目と共通受検者を同時に利用した等化

ところで、通常 IRT による等化では、通常共通項目による等化か共通受検者による等化のいずれかが実現可能となるように、テストフォーム間に共通する問題項目や受検者群の配置を設計する。一方、多次元 IRT の応用場面のひとつとして、それまで 1 次元の IRT を利用して運用されてきた複数のテストを、多次元 IRT の枠組みで統合するケースが考えられるが、この場合にはテストを構成するすべての項目を共通項目、すべての受検者を共通受検者と見なし、従来の 1 次元の IRT による項目パラメータ、受検者パラメータを等化先として形式的に等化の手続きを実行することができる。あるいは多次元 IRT での運用を前提に新規に開発されたテストの場合にも、測定内容の安定性を重視する立場から、一旦 1 次元の IRT を利用して各次元に対応する項目特性値や受検者特性値を推定し、これをターゲットとして形式的に等化の手続きを利用することができる。加えて、多次元 IRT の通常の運用場面においても、共通項目と共通受検者の両方を活用する等化デザインを考えることは可能である (e.g. Davey *et al.*, 1996)。したがって、多次元 IRT の応用場面では、共通項目と共通受検者の両方の情報を同時に利用可能なケースが存在すると考えられるが、その一方で 2 つの情報が同時に利用可能な状況に対応した多次元 IRT の等化法は現時点では提案されていない。

共通項目と共通受検者の両方の情報を利用することは、広い意味で精度が高くバランスのよい等化結果につながる可能性がある。共通項目による等化と共通受検者による等化を統合することで 2 つの情報を同時に考慮した新たな等化法を提案し、この方法によって有効な等化結果が得られることを確認することが、本研究のもうひとつの目的である。

7.2 本研究の結果と意義

共通受検者による等化法の提案

共通受検者デザインの場合の主な等化法である Hirsch の方法について、等化先と等化後の受検者特性値の近接性が不十分である点、等化後の相関構造が等化先のそれと一致しない点を課題として指摘した。受検者特性値の近接性を高めるためには、等化に関わる制約のより少ない方法が有効であると考えられる。一方、相関構造を再現するためには、この構造を制約として考慮した等化法が求められる。単一の方法で、これら両方を同時に追求することは難

しいと考えられるため、本研究では前者に対応するものとして共通受検者に注目した斜交プロクラステス回転による等化法を、後者に対応するものとして相関構造に制約のある共通受検者に注目した重み付きプロクラステス回転による等化法を案出した。多次元 IRT の既存の等化法の多くは、近似解もしくは数値的な解法が知られているのみであるが、ここで提案した 2 つの等化法については、共に厳密解 (解析解) が得られている。

また、上記の方法の特性を明らかにするために、先行研究で示された方法と比較する形で計算機による実験を行った。実験の結果、共通受検者に注目した斜交プロクラステス回転による方法が、等化先と等化後の受検者特性値の近接性という点で、他の方法と比較して良好な結果を与えることが確認できた。また、相関構造に制約のある共通受検者に注目した重み付きプロクラステス回転による方法では、想定どおり等化先と等化後の相関構造が一致すること、等化先と等化後の受検者特性値の近接性という面でも、既存の方法である Hirsch の方法と同等か、特に次元間の相関係数が大きい場合にはやや有利な結果が得られる可能性が高いことが確認できた。なお、実際のデータである PISA データを利用した分析についても、同様の傾向を確認した。

ところで、やはりシミュレーション研究の結果から、共通受検者に注目した斜交プロクラステス回転による方法は、特に問題項目数が少ない場合に項目パラメータの等化結果が不安定になる傾向があること、受検者特性値の次元間相関の構造が等化先と大きく異なったものとなる場合があることが把握できた。これらの結果を総合すると、共通受検者による等化を行う場合には、重み付きプロクラステス回転による方法が総合的にみて有効な手法であるものと思われる。

さて、この実験の結果から得られた別の観点での示唆として、等化先と等化後の受検者特性を一致させるという点で有効な等化法が、必ずしも項目特性値についても良好な結果をもたらすわけではないという知見がある。多次元 IRT の応用場面では、前述の通り共通受検者と共通項目が共に存在する場合があるが、上記のことは既存の等化法を利用した場合、等化の結果は共通受検者による等化、共通項目による等化のどちらを用いるかによって異なったものとなり、選択した等化法の観点に偏った傾向を有することを示唆するものである。何らかの方法で偏りを回避したバランスのよい等化が実現できることが望ましいと考えられる。本研究ではこの問題を解決する手法として、共通受検者と共通項目の情報の両方を利用することで等化結果の安定性を高めるような等化法の可能性について検討した。

共通項目と共通受検者を同時に利用した等化法の提案

共通受検者による重み付きプロクラステス回転による等化法に、共通項目による重み付きプロクラステス回転法を統合することで、共通項目と共通受

検者の両方の情報を考慮した新たな等化法を提案した。2つの等化法の統合にあたっては、それぞれの等化基準の重み付きの線形結合を等化基準として設定し、これを等化先と等化後の相関係数が一致するという制約条件下で最小化するというアプローチを採用した。重みの値によって共通受検者による基準と共通項目による基準のどちらを重視するかを考慮することができるため、提案手法は本研究でそれぞれ新規に提案した共通受検者による重み付きプロクラステス回転による等化法と、共通項目による重み付きプロクラステス回転による等化法を、さらに一般化した手法であると捕らえることができる。なおこの方法の場合は、解は数値的な計算によって求めることとなる。

この手法についても、計算機実験を行って特性を確認した。結果、この場合にも、想定どおり等化先と等化後の相関構造が一致することが検証できた。また、重みを中間的な値として共通項目と共通受検者の両方の情報を利用した場合には、いずれかの基準に偏らないバランスのよい等化結果が得られることが確認できた。本研究で提案した共通項目と共通受検者の両方の情報を同時に利用する等化法はしたがって、等化結果の安定性という面に特徴のある等化法であり、広い意味での等化精度の向上に有効な方法であると考えられる。

各手法の実行条件と適用範囲

各手法の実行条件と適用範囲についてまとめる。まず、受検者特性値に注目した斜交プロクラステス回転法と、同じく受検者特性値に注目した相関構造に制約のある重み付きプロクラステス回転は、ともに共通受検者デザインの等化の場合に対応した方法である。したがって、共通受検者が存在することが、これらの手法を利用する場合の前提となる。なお、いずれの方法についても、1次元のIRTを利用して推定した複数の尺度のスコアをターゲットとして等化を行うことが可能である。

共通項目に注目した重み付きプロクラステス回転法は、共通項目デザインに対応した方法である。この方法についても受検者特性値の次元間相関を制約として考えるため、等化先の次元間相関の相関の情報が必要となる。ただし、等化先の次元間相関の値は、実際のデータから計算した値である必要性はなく、たとえば過去の研究の蓄積などからターゲットとなる相関構造が明らかかな場合には、その情報を利用することも可能である。

共通受検者と共通項目の両方を利用した重み付きプロクラステス回転法は、共通受検者と共通項目が混合したデザインに対応した方法であり、両方の情報を利用する。したがって、たとえばそれまで1次元のIRTで運用されてきたテストを、多次元IRTの枠組みで統合するなど、共通項目と共通受検者がともに存在するような状況での利用が想定される。

なお、本研究では多次元IRTのモデルとして、2パラメータ・ロジスティックモデルを想定して議論を進めたが、上記の各等化法は項目パラメータと受検

者特性値が、受検者特性値を θ 、項目 j のパラメータを a_j, d_j として、 $a_j\theta + d_j$ のように因子分析モデルの形をしているモデルであれば、広く利用することが可能である。本研究で紹介した中では、たとえば多次元段階反応モデルがこれにあたる。他に、1次元のIRTの場合の1パラメータ・ロジスティックモデルや3パラメータ・ロジスティックモデルを多次元化した補償型モデルが、適用範囲となる。

また本研究は、因子分析におけるターゲットの存在する回転法であるプロクラステス回転を、多次元IRTの文脈で捉えなおしたものとして位置づけることが可能である。その意味で、本研究で提案した各等化法は、そのまま因子分析の場合の回転法として利用することもできる。

全体的な意義

前述のとおり、多次元IRTのテストの採点への活用はあまり進んではおらず、この背景に等化法に関する研究の不足があることが指摘されている。特に共通受検者による等化法は、現時点ではほとんど提案されておらず、研究が急がれる領域である。

本研究では、多次元IRTの共通受検者を利用した等化法として、斜交プロクラステス回転法、相関構造に制約を課した重み付きプロクラステス回転による方法、共通受検者に加えて共通項目の情報を考慮した方法の3つの方法を新たに提案した。これらの方法は、それぞれ既存の等化法にはない特徴を有しているため、多次元IRTの活用範囲の拡大に資する可能性が高いものと考えられる。特に、重み付きプロクラステス回転に基づく方法は、等化先と等化後のテストの相関構造が一致する特性がある。本来同じ特性群を測定することを意図された複数のテストフォームにおいて、次元間の相関構造が異なっている場合、次元ごとのスコアの重み付き合計点の解釈や、次元間のフィッシャー情報量の比較などで不都合を生じる可能性がある。重み付きプロクラステス回転の相関構造を一致させるという特徴は、この点においても実用上の有効性が高いものと思われる。

7.3 今後の課題

本研究の課題と今後の期待される研究についてまとめる。

実験の要因と手法

本研究の課題として、まず計算機実験の要因に関する事柄が指摘できる。実験では、受検者特性値の次元間の相関の大きさ、項目数、次元数などを要因として取り上げたが、多次元IRTの等化では他にも、受検者数、 α パラメータの

構造、パラメータの推定アルゴリズムなど、1次元のIRTの場合に比較して非常に多くの要素が相互に組み合わさる形で結果に影響するものと考えられる。また、次元数が3以上の場合には、相関係数の大きさに加えて、どの次元間の相関が大きく、どの次元間の相関が小さいかという相関係数の構造も等化結果に影響を及ぼす可能性がある。共通項目と共通受検者の両方の情報を利用する等化法では、これらを結合する重み係数の設定が必要となるが、この重みの設定方法について何らかの目安を得るためにも、多様な設定下での実験が必要であると思われる。今後さらに幅広い要因を考慮した研究を行い、等化法の性質を調べていく必要があるものと考えている。

また、本研究ではPISAの実際のデータを等化した結果について報告したが、このような実際のデータを利用した場合についても、今後様々なテストを分析することで、一般的な傾向を把握していく必要があるものといえる。

等化基準

つぎの課題として、本研究で提案した手法はいずれも、共通受検者や共通項目の特性値のユークリッド距離を直接に最小化することを目指したものである点が指摘できる。行列間の距離には他にも様々な定義があり得る(e.g. Gower & Dijksterhuis, 2004)。また、1次元の場合にはStocking & Lordの方法やHaebara法など特性値の直接の距離ではなく、項目特性曲線のギャップの最小化という形で等化を定式化する方法が存在する。多次元IRTにおいても、TCF法やICF法など同様のコンセプトに基づく方法が開発されている。

たとえば、TCF法の枠組みの中で、相関構造に制約を設定した等化を行うなど、これらの基準について提案手法を拡張する方法を考えることができるが、本研究ではそのような方向については十分検討できなかった。この点についても、今後の課題であると考えている。

最後に、本研究をさらに推し進める場合の今後の展開の方向性についてまとめる。

共通項目と共通受検者の両方を利用した他の等化法

本研究では、共通項目と共通受検者の両方の情報を考慮した等化法として、共通項目に注目した重み付きプロクラステス回転法と共通受検者に注目した重み付きプロクラステス回転法を統合するアプローチを提案したが、この方略は先行研究の方法を含めた他の等化法にも応用可能なものである。たとえば、共通受検者による等化法であるHirschの方法や、共通項目によるLi & Lissitzの方法を拡張して、共通受検者と共通項目の両方を同時に考慮する等化法をそれぞれ構成することができる。また、本研究において共通受検者による等化法

として提案した斜交プロクラステス回転に基づく方法についても、これを拡張することで共通項目の情報を併せて考慮する等化法が実現できると考えられる。特に、共通受検者に注目した斜交プロクラステス回転による方法では、項目数が少ないときに共通項目の等化後の値が等化先に十分に近接できない問題があった。これは、斜交プロクラステス回転による方法が比較的小さい制約条件下で等化を実現する方法であり、そのためにターゲットである等化先の値への追従性が高くなりすぎる一種の過剰適合の状態が生じ、共通受検者をターゲットとした場合にはもう一方の共通項目側の追従性が却って阻害されたものと解釈できる。たとえば、斜交プロクラステス回転について、共通項目と共通受検者の情報を同時に利用する方向にこれを拡張することで、上記の問題を抑制した安定性の高い等化が実現できる可能性がある。この方法については、本研究の延長線上の研究として検討すべきテーマのひとつであると考えている。

ところで、2パラメータ・ロジスティックモデルなどの1次元のIRTのモデルは、多次元IRTの補償型モデルに包含されるものとして捉えることができた。現状、テストの運用場面では1次元のIRTが主流であるが、本研究で提案した等化法は、次元数が2以上の場合に固有の方法となっている¹。ただし、共通項目と共通受検者の両方の情報を利用するというコンセプト自体は1次元のIRTの場合にも応用できる可能性がある。

共通項目による等化を行った場合にはもう一方の受検者の、共通受検者による等化を行った場合には項目側のパラメータの追従性が却って低くなるという問題は、1次元のIRTの等化の場合にも同様に生じる可能性がある。また、1次元のIRTを想定したテストについても、共通項目と共通受検者を共に含む等化のデザインを考えることは可能である。このような場合には通常、共時等化などの等化法が利用されるが、本研究で提案したような、項目と受検者の両方をターゲットとした等化法も有効な手法の一つである可能性がある。本研究の方法を1次元の場合に拡張し、この点を確認する方向の研究についても検討の価値があるものと考えられる。

また、1次元のIRTで運用されている複数のテストについて、テスト間の相関構造を維持したまま、それぞれ等化することが要請される場面が実務上想定できる。この場合には、必ずしも多次元IRTモデルによるパラメータ推定を経る必要はないが、等化の手続きは、本研究の延長線上にあると思われるため、具体的な手順について示すことには意味があるものと考えている。

分散制約の緩和

本研究では、主に相関構造に制約を課した多次元IRTの等化法について検討した。テストによって測定される受検者特性値の次元間の相関構造は、各次

¹たとえば、次元が1つの場合には相関構造を一致させるという制約には意味がない

元のスコアが何を表しているかということと関連した重要な指標であると考えられる。たとえば、本来同じ内容を測定することを狙いとした2つのテストフォームが存在する場合、これらの相関構造が異なっていることは、スコアの解釈の面において不都合を生じる。また、共通受検者の特性値の分散についても、等化先と等化後では同じ値であると考えることには一定の合理性がある。したがって、本研究では等化後の相関と分散がそれぞれ等化先のテストフォームのそれと一致するという制約下で等化を行うことに力点をおいて、いくつかの手法を提案した。

一方、先行研究においてこれらの条件がどのように捉えられているかを概観すると、たとえば、共通受検者に基づく Hirsch の方法では、潜在特性の基底を一致させるというプロセスにおいて分散と相関の一致制約を一部考慮しているように思われる。しかし、Min の方法などの共通項目に基づく方法では、基本的にこれらの条件については考慮されておらず、共通項目による等化法と共通受検者による等化法で対応が分かれているように思われる²。

さて、分散と相関構造にそれぞれ制約を置くことは、分散共分散行列の値を制約することと同値であり、本研究でもこの方針で新たな等化法の提案を行った。ただし、分散と相関に関する制約は、本来別々に取り扱うことが可能である。具体的には、たとえば共通受検者に注目した重み付きプロクラステス回転法において、分散には制約をおかず、相関だけを指定された構造に一致させるような等化法を開発することは可能である³。等化に関する制約を緩和することは、等化先の実際のデータに対する追従性を高める方向に寄与すると考えられる。等化後の特性値を実際のデータにできるだけ近づけたいという動機が強い場合には、このような方法が求められる場面もあり得るものと想定されるため、提案手法について、分散制約を緩和した等化法を用意しておくことには意味があるものと考えられる。

等化先の項目特性値が一部しか指定できない場合

因子分析の文脈では、回転のターゲットとなる因子負荷量が一部しか指定できない場合の回転法が知られている。このような回転法は不完全プロクラステス回転法 (Browne,1972a; Browne,1972b; Browne,2001) と呼ばれる。たとえば、2 因子構造を想定して開発された尺度において、一部の項目群については第 1 因子への因子負荷量が 0 に近く、別の項目群については第 2 因子への負荷量が 0 に近いことが事前の想定から明らかであるが、もう一方の因子への負荷量については不明である場合には、このような手法が有効である。多次元 IRT においても、各問題項目と潜在特性との関連性については何らかの

²ただし、本研究では共通項目と共通受検者の情報を共に利用する等化法を提案する過程で、分散共分散行列に制約のある場合の共通項目に注目した等化法を導出した。

³共通受検者による等化において、分散だけに制約を課した場合は、因子スコアによる直交もしくは斜交プロクラステス回転となる

事前の仮説が存在することが稀ではない。具体的には、問題項目があらかじめいずれかの次元を想定して作成されている場合や、特定の次元とは問題項目の性質上とは無関係であることが明らかである場合などが考えられる。また、本研究で指摘したとおり、受検者特性値の相関についても、事前に構造を想定することが適当であるケースが多いように思われる。したがって、潜在特性の相関構造を指定した不完全プロクラステス回転法が開発できれば、実際のテストの運用ばめんにおける有用性が高いものと考えられる。このような方法は、本研究で示した相関構造を指定した等化法を一部の項目について項目特性値にターゲットが存在する場合に拡張することで実現できる。この手法についても、今後研究を進める意義があるものと思われる。

さて、多次元 IRT の等化法については、現状提案されているもの以外にも様々な可能性があり得る。また、各種の応用場面に適した等化法を提案し、その特性を見極めていくことは、多次元 IRT モデルの普及の鍵になると考えられる。本研究で提案した方法を含めて、さらなる検証と展開が望まれる領域であり、今後もこれに貢献していきたいと考えている。

付録

付録 1 :次元間相関の合成スコアに対する影響

個人毎の総合スコアとして、複数の尺度の重み付き和が利用されることがよくある。この方法は、多次元 IRT の文脈では受検者特性値の次元に対する加重和に対応する。ここでは、次元間の相関構造が合成スコアに影響を与えることの例を示す。

次元数が 3 の場合に対応した、つぎの受検者特性値の次元間の相関係数行列 R_1 と R_2 について考える。

$$R_1 = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 1.0 & 0.3 & 0.0 \\ 0.3 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (7.1)$$

R_1 は次元間が無相関の場合であり、 R_2 は R_1 について第 1 次元と第 2 次元の間の相関を 0.3 に置き換えたものである。テストの次元間の相関構造が、 R_1 に従う場合と R_2 に従う場合で個人毎の合成スコアが一致しないことを、計算機実験によって調べる。

まず、 $n = 10,000$ として、 R_1 に従う各次元の平均が 0 の正規分布から n 人分のデータを乱数で生成した。各次元の分散は 1.0 とする。したがって R_1 は分散共分散行列と見なすことができる。生成したデータを X_1 とする。 X_1 は行数が n 、列数が 3 の行列である。このデータについて

$$\mathbf{y}_1 = X_1 \mathbf{1} \quad (7.2)$$

として、 \mathbf{y}_1 を求めた。 $\mathbf{1}$ は 1 を縦に 3 つ並べたベクトルである。 \mathbf{y}_1 は個人毎に次元間の和を求めたもので合成スコアに相当する (各次元の重みは 1.0 とした)。

つぎに、 X_1 の実現値の分散共分散行列 \bar{R}_1 および R_2 のスペクトル分解を $\bar{R}_1 = P\Lambda P^T$, $Q\Delta Q^T$ として、

$$X_2 = X_1 P \Lambda^{-1/2} \Delta^{1/2} Q \quad (7.3)$$

によって、 X_1 を相関構造が R_2 となるように変換した。ただし、 $PP^T = QQ^T = I$ で、 Λ と Δ はそれぞれ R_1 と R_2 の固有値を対角にならべた行列である⁴。 X_2

⁴実現値の相関行列 \bar{R}_1 の代りに R_1 を利用することも考えられるが n が十分に大きい場合

についても

$$\mathbf{y}_2 = X_2 \mathbf{1} \quad (7.4)$$

として合成スコアを求めた。

図(7.1)は、横軸を \mathbf{y}_1 、縦軸を \mathbf{y}_2 として n 人の受検者をプロットした散布図である。 \mathbf{y}_1 と \mathbf{y}_2 の相関係数は .789 であり、図からも明らかに \mathbf{y}_1 と \mathbf{y}_2 は

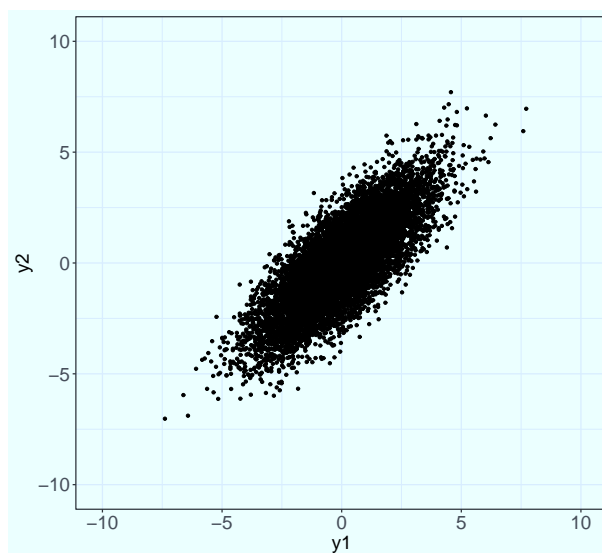


図 7.1: 相関構造が異なる 2 つの合成スコアの散布図

一致していない。また、スピアマンの順位相関は .774 であり受検者の順位についても相関構造の影響を受けていることがわかる。

以上の例から、テストの次元間の相関構造は対応する合成スコアに影響を与えており、同じ内容を測定するテストフォームであっても、相関構造が異なる場合には総合スコアによる受検者の順位などが一致したものとはならないことが理解できる。

上の例で、 \mathbf{y}_2 は

$$\mathbf{y}_2 = X_1 P \Lambda^{-1/2} \Delta^{1/2} Q \mathbf{1} \quad (7.5)$$

と書けるが、これは X_1 の各次元の値に対する $P \Lambda^{-1/2} \Delta^{1/2} Q \mathbf{1}$ による重み付き和として解釈できる。 $\bar{R}_1 \neq R_2$ の場合、

$$P \Lambda^{-1/2} \Delta^{1/2} Q \neq I \quad (7.6)$$

であるから、上の例の場合に限らず、一般にテストの相関構造は合成スコアに影響を及ぼすものと考えられる。

には、結果はほとんど変わらない。

付録 2 : 受検者特性値 (因子スコア) に注目した直交プロクラステス回転

次の問題について考える.

$$\min. f = \frac{1}{2} \text{tr}(\Theta - \Theta_g W)^T (\Theta - \Theta_g W) \quad (7.7)$$

$$\text{s.t. } \text{diag}(W^T S_g W) = I \quad (7.8)$$

いま W が $W^T W = I$ を満たす直交回転行列であるとする. 上式の f を展開すると,

$$\begin{aligned} f &= \frac{1}{2} \text{tr}(\Theta^T \Theta - 2\Theta^T \Theta_g W + W^T \Theta_g^T \Theta_g W) \\ &= \frac{1}{2} \text{tr}(nS - 2\Theta^T \Theta_g W + nW^T S_g W) \end{aligned}$$

となるが, $W^T W = I$ より $\text{tr}(W^T S_g W) = \text{tr}(S_g)$ であるから, (7.7) 式の最小化は $\Theta^T \Theta_g W$ の最大化と同値である. $\Theta^T \Theta_g$ が正則な場合, このような W は

$$W = \Theta^T \Theta_g (\Theta_g^T \Theta \Theta_g)^{-1/2}$$

と求めることができる (e.g. 市川, 2010). さて, この場合の Θ_g の変換後の分散は $W^T S_g W$ の対角成分であるが, この各成分の値は一般には 1 にならない. ただし特に $S_g = I$ の場合には $W^T W = I$ より

$$W^T S_g W = W^T W = I$$

となる. $S_g \neq I$ の場合には変換後の Θ_g の分散で W を調整した

$$W^* = W [\text{diag}(W^T S_g W)]^{-1/2}$$

を利用する.

付録 3 : 受検者特性値 (因子スコア) に注目した斜交プロクラステス回転

いま f が W の関数であることを明示的に $f(W)$ と書くことにすると, (7.7) 式の (7.8) 式による条件付き最小化問題は

$$g = f(W) - \frac{1}{2} \text{tr} [\Gamma (\text{diag}(W^T S_g W) - I)] \quad (7.9)$$

の最小化と同値であり、解は

$$\begin{cases} \frac{\partial g}{\partial W} = \frac{\partial f(W)}{\partial W} - S_g W \Gamma = 0 & (7.10) \\ \frac{\partial g}{\partial \Gamma} = \text{diag}(W^T S_g W) - I = 0 & (7.11) \end{cases}$$

を満たす。ただし、 Γ はラグランジュの未定乗数を表す対角行列である。ここで、

$$G = \frac{\partial f(W)}{\partial W}$$

とすると、(7.10) 式から

$$G = S_g W \Gamma$$

であり、上式の両辺に左から W^T をかけると、

$$W^T G = W^T S_g W \Gamma$$

となる。ここで Γ が対角行列であることに注意すると、任意の j ($j = 1, 2, \dots, m$) について

$$(W^T G)_{jj} = (W^T S_g W)_{jj} \Gamma_{jj}$$

が成り立つ。ただし上式の下付きの添字 jj は各行列の j 行 j 列成分を表す。上の関係を考慮すると、(7.11) 式より

$$\Gamma = \text{diag}(W^T G)$$

であることがわかる。これを (7.10) 式の Γ に代入すると

$$\begin{aligned} & \frac{\partial f(W)}{\partial W} - S_g W \text{diag}(W^T G) \\ &= n S_g W - \Theta_g^T \Theta - S_g W \text{diag}[W^T (n S_g W - \Theta_g^T \Theta)] \\ &= S_g W \text{diag}(W^T \Theta_g^T \Theta) - \Theta_g^T \Theta \end{aligned}$$

(途中 (7.11) 式を利用した) となるから、

$$S_g W \text{diag}(W^T \Theta_g^T \Theta) - \Theta_g^T \Theta = 0$$

であり、 W に関連する項を右辺にまとめて

$$S_g^{-1} \Theta_g^T \Theta = W \text{diag}(W^T \Theta_g^T \Theta) \quad (7.12)$$

を得る。いま、

$$V = S_g^{-1} \Theta_g^T \Theta$$

とすると, (7.12) 式から

$$V = W \text{diag}(W^T S_g V) \quad .$$

さらに, $[\text{diag}(W^T S_g V)]_{jj} \neq 0, (j = 1, 2, \dots, m)$ を仮定して⁵

$$K = [\text{diag}(W^T S_g V)]^{-1} \quad (7.13)$$

とすると,

$$W = VK \quad .$$

これを, (7.13) 式に再び代入すると

$$K = [\text{diag}(K^T V^T S_g V)]^{-1} = [\text{diag}(V^T S_g V)]^{-1} K^{-1}$$

となる. 上式を K について解くと

$$K = [\text{diag}(V^T S_g V)]^{-1/2}$$

であるから, 結局 W は

$$W = VK = V[\text{diag}(V^T S_g V)]^{-1/2}$$

となる⁶ (以上).

⁵ $[\text{diag}(W^T S_g V)]_{jj} = 0$ となるのは, W の j 列の各成分が全て 0 の場合か, $\Theta_g W$ すなわち等化後の能力特性値と等化先の能力特性値 Θ の第 j 次元が直交する場合であるが, 前者は (7.11) 式に反する. また, 後者の場合は等化の設定に問題があると考えられる.

⁶上の W は本文の W_g^* に対応する

謝辞

本論文は、筆者が東京工業大学大学院 社会理工学研究科 人間行動システム専攻 博士課程 において行った研究をまとめたものです。

本論文の執筆にあたり、終始暖かい激励とご指導、ご鞭撻を頂いた指導教官である東京工業大学リベラルアーツ研究教育院 前川眞一教授に心より感謝申し上げます。また、本論文をご精読頂き有用なコメントを頂きました東京工業大学リベラルアーツ研究教育院 林 直亨 教授, 室田 真男 教授, 松田 稔樹 准教授, 東京工業大学工学院 中山 実 教授に深く感謝致します。

芝祐順 東京大学名誉教授 には、進学して研究することを勧めていただきました。それから随分と時間がたってしまいました。ようやくアドバイスに従うことができました。

前川研究室のゼミ生, 卒業生の皆様にも研究にあたって多くのご支援をいただきました。なにより研究を進めるにあたって、様々な面でその存在が刺激となりました。感謝申し上げます。

最後に、執筆を支援してくれた妻にも感謝を伝えたいと思います。

関連する研究

第 2 章

沖嘉訓・前川眞一 (2014). 多次元項目反応モデルによるテストデータの分析. 日本テスト学会第 12 回大会 発表論文抄録集.

第 4 章

沖嘉訓・前川眞一 (2013). 多次元項目反応理論における共通受検者の能力特性値による等化について. 日本テスト学会第 11 回大会 発表論文抄録集.

沖嘉訓・前川眞一 (2015). 多次元項目反応モデルにおけるプロクラステス回転を応用した共通受検者による等化法の比較. 行動計量学, 第 42 巻第 2 号,117-128.

第 5 章

沖嘉訓・前川眞一 (2017). 多次元項目反応モデルにおける共通項目と共通受検者の両方を同時に考慮した等化法の提案. 行動計量学 (in press).

Yoshinori Oki & Shin-ichi Mayekawa. (2016). Equating multidimensional IRT parameters when both common items and common persons are available. International Conference on Computational Statistics (oral-presentation).

参考文献

- [1] Ackerman, T.A. & Gierl, M.J., & Walker, C.M.(2005). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*. 22,3,37-51.
- [2] Arai,S., & Mayekawa,S.(2011). A comparison of equating methods and linking designs for developing an item pool under item response theory. *Behaviormetrika*, 38,1-16.
- [3] 荒井清佳・菊池賢一・前川眞一. (2013). 三つ以上のテストフォームを等化する場合の等化精度の違いについて. *日本テスト学会第11回大会発表論文抄録集*, 60-61.
- [4] Bock,R.D. (1972).Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*,37,29-51.
- [5] Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement.*, 12,3,261-280.
- [6] Brown,A., & Maydeu-Olivares,A.(2010).Item response modeling of paired comparison and ranking data.*Multivariate Behavioral Research.*,45,6,935-974.
- [7] Brown,A., & Maydeu-Olivares,A.(2012). How IRT can solve problems of ipsative data in forced-choice questionnaires. *American Psychological Association*.
- [8] Browne,M.W. (1972a). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*. 25,115-120.
- [9] Browne, M.W. (1972b). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*. 25, 207-212.
- [10] Browne, M.W.(2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*. 36,111-150.

- [11] Chalmers,R.P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of statistical software*,48,6.
- [12] Davey,T. , Oshima,T.C., & Lee, K. (1996). Linking multidimensional item calibrations. *Applied Psychological Measurement*. 405-416.
- [13] 藤森進 (1985) 項目反応モデルに基づく多次元新モデルの提案. *東京工業大学人文論叢*, No,11,79-89. 20,4,
- [14] 藤森進 (1988). OR 型多次元項目反応モデルによる TPI 尺度データの分析. *教育心理学研究*,36,96-103.
- [15] Gower, J.C., & Dijksterhuis,G.B. (2004). *Procrustes problems*. Oxford university press.
- [16] Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22, 144-149.
- [17] Hirsch,T. (1989).Multidimensional equating. *Journal of educational measurement*.26,4,337-349.
- [18] 市川雅教 (2010). *因子分析 行動計量の科学 7*. 朝倉書店.
- [19] 池田央 (2001). *項目応答理論. 現代テスト理論*, 朝倉書店.
- [20] 泉毅・山野井真児・山田 剛史・金森保智・対馬英樹, (2012). 共通項目数が等化の精度に及ぼす影響：大規模学力テストデータを用いた探索的研究. *教育実践学論集*,13,49-57.
- [21] Jennrich,.R.I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika*. 66,289-306.
- [22] Jonathan,P.W. (2014). Multidimensional test linking. In Reise, S.P., & Revicki, D.A. (Eds.). *Handbook of item response theory modeling*.
- [23] 加藤健太郎・山田剛史・川端一光 (2014). *R による項目反応理論*. オーム社.
- [24] Koschat, M.A., & Swayne, D.F. (1991). A weighted Procrustes criterion. *Psychometrika*. 56.2.229-239.
- [25] 高大接続改革プロジェクトチーム (2016). *高大接続システム改革会議「最終報告」*.
- [26] Li, Y. H., & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied psychological measurement*, 24,115-138.

- [27] Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum.
- [28] Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- [29] Min, K. (2003). The impact of scale dilation on the quality of the linking of multidimensional item response theory calibrations. PhD thesis, Michigan State University.
- [30] Min, K. (2007). Evaluation of linking methods for multidimensional IRT calibrations. *Asia Pacific Education Review*, 8, 1, 41-55.
- [31] 光永悠彦・前川眞一 (2013). 多群 IRT モデルにおけるモデル簡素化の評価 -水平等化場面のシミュレーションを通じて-. *行動計量学*, 40, 29-42.
- [32] 光永悠彦・御園真史・水町龍一 (2013). 共通項目デザインを用いた多値型データによるテストにおける等化方法の比較 -大学1年次向け数学コンピテンシーテストのデータを用いて-. *日本テスト学会第11回大会発表論文抄録集*, 56-59.
- [33] 村木英治 (2011). 項目反応理論. 朝倉書店.
- [34] Muraki, E., & Carlson, J.E. (1995) Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, 19, 73-90.
- [35] 村山航 (2012). 妥当性概念の歴史的変遷と心理測定学的観点からの考察. *教育心理学年報*, 51, 118-130.
- [36] 日本テスト学会 (編). (2007). テスト・スタンダード 日本のテストの将来に向けて. 金子書房.
- [37] 仁田善雄・前川眞一・柳本武美・前田忠彦・吉田素文・奈良信雄・石田達樹・福島統・齋藤宣彦・福田康一郎・高久史磨・麻生武志 (2005). 項目反応理論を用いた第1回共用試験医学系 CBT の統計解析. *医学教育*, 36(1), 3-9.
- [38] 岡田謙介・前川眞一 (2014). 補償型・非補償型を包含する多次元項目反応理論モデル. *日本テスト学会第12回大会発表論文抄録集*.
- [39] 沖嘉訓・前川眞一 (2013). 多次元項目反応理論における共通受検者の能力特性値による等化について. *日本テスト学会第11回大会 発表論文抄録集*.

- [40] 多次元項目反応モデルによるテストデータの分析. 日本テスト学会第12回大会 発表論文抄録集.
- [41] 沖嘉訓・前川真一 (2015). 多次元項目反応モデルにおけるプロクラステスト回転を応用した共通受検者による等化法の比較, 行動計量学, 42-2, 117-128.
- [42] 沖嘉訓・前川真一 (2017). 多次元項目反応モデルにおける共通項目と共通受検者の両方を同時に考慮した等化法の提案. 行動計量学 (in press).
- [43] Oshima,T.C., Davey,T.D., & Lee,K. (2000). Multidimensional linking; four practical approaches. *Journal of educational measurement*, 37, 357-373.
- [44] R Core Team (2015). R: a language and environment for statistical computing. <http://www.R-project.org/>.
- [45] Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied psychological measurement*, 9, 401-412.
- [46] Reckase, M.D. (2009). *Multidimensional item response theory*. Springer.
- [47] Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied psychological measurement*, 15,361-373.
- [48] Reise,S , Moore,T. & Maydeu-Olivares,A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models. *Educational and Psychological Measurement*,71,4,684-711.
- [49] Reise, S.P., & Revicki, D. A. (2014). *Handbook of item response theory modeling: Applications to typical performance assessment*. Routledge.
- [50] Rupp, A.A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
- [51] 芝祐順.(1979). 因子分析法 (第2版). 東京大学出版会.
- [52] 芝祐順 (編) (1991). 項目反応理論—基礎と応用. 東京大学出版会.
- [53] 静哲人 (2007). 基礎から深く理解するラッシュモデリング—項目応答理論とは似て非なる測定のパラダイム. 関西大学出版部.

- [54] Simon, M.K. (2008). Comparison of concurrent and separate multidimensional IRT linking of item parameters. Ph.D. thesis, University of Minnesota.
- [55] 孫媛 (1997). 多次元データに対する項目反応モデル. 学術情報センター紀要 第9号, 103-111.
- [56] Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- [57] Takane, Y., & Leeuw, J.D. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 3, 393-408.
- [58] van der Linden, W.J., & Hambleton, R.K. (Ed.) (1997). *Handbook of modern item response theory*. Springer.
- [59] 渡辺 直登・野口 裕之 (1999). *組織心理測定論—項目反応理論のフロンティア*. 白桃書房.
- [60] 柳井晴夫・繁栞算男・前川眞一・市川雅教 (1990). *因子分析-その理論と方法*. 朝倉書店.
- [61] Yao, L., & Boughton, K. (2009). Multidimensional linking for tests with mixed item types. *Journal of Educational Measurement*, 46, 2, 177-197.
- [62] Yon, H. (2006). *Multidimensional item response theory (MIRT) approaches to vertical scaling*. Ph.D. thesis, Michigan state university.