流通科学大学論集－経済・経営情報編－第 17 巻第 1 号, 9-33(2008)

# テストの正誤反応行列から得られる次元性情報に関する考察

# Test Dimensionality Defined by the Binary Response Matrix

平越裕之*・井澤廣行**
Hiroyuki Hirakoshi and Hiroyuki Izawa

This paper investigates the relationship between the variation of the first contribution-ratios in the principal components analyses of the binary response matrices and the distributional characteristics of the parameter estimates on the Rasch measurement, especially, the sizes of their standard deviations. The notable findings concerning the components analyses of the item responses are as follows: 1) The smaller first contribution-ratios are related to the larger standard deviations of the difficulty estimates and 2) the larger first contribution-ratios are related to the larger standard deviations of the ability estimates. Those concerning the components analyses of the examinee responses are symmetrical.

**Key words**: Binary responses, Principal components analysis, First contribution-ratio, Item response theory, Rasch item analysis model.

## I．はじめに

テストの各設問項目に対する各受験者の正誤反応によって、受験者能力母数と項目困難度母数を含めた項目属性を分析しようとした近代テスト理論が一次元項目反応理論(Lord and Novick, 1968, cited in Reckase[1], 1979, p. 207)である。項目反応理論では、受験者能力と項目困難度の母数(parameter)のみから成る一母数モデルに、項目属性として「項目弁別力」を追加想定したものを二母数モデル、更に、「当て推量母数」を付加したものを三母数モデルと呼称している(静[2]、2007、pp. 345- 365、参照)。数式の上では、項目反応理論一母数モデルは Rasch 項目分析モデル[3](Rasch, 1960)（以降 Rasch モデル）と同一である。然しながら、母数が一つの単純なモデルと位置付けている項目反応理論体系に比べ、十分統計量(sufficient statistics)の存在に基づく母数分離性(parameter separability)に着眼した(Rasch[3], 1960, pp. 174-178)Rasch モデル体系では、Wright[5] (1992)を含める Rasch モデル識者によって、その測定理論妥当性(Rasch[4], 1961, p. 326)が強く意識されている。

一次元項目反応理論の適用が妥当であるかの判断に際しては、各受験者と各項目を行と列に配置した正誤反応行列から、各項目間行列を導出してこの相関行列を基に主成分分析を行った上で、第一主成分(固有値)寄与率の大小を判断する方法が一般的である(芝[6]、1991、p. 169； 渡辺・野口[7]、1999、

* 流通科学大学情報学部、〒651-2188 神戸市西区学園西町3-1
** 流通科学大学サービス産業学部、〒651-2188 神戸市西区学園西町3-1

(2008年 4 月 1 日受理)

p. 23、参照)。なお、因子分析における第一因子寄与率は、常に主成分分析第一主成分寄与率以下であるため、本稿では主成分分析の第一主成分寄与率のみで議論を進める。重要事項として、その許容下限寄与率は、Reckase[1](1979, p. 227)と豊田[8](2002、p. 30)によれば、20%であるとされている。一方、Raschモデルにおいては、正誤反応データへのモデル適用自体の是非に関する制限は存在せず、むしろ、順序尺度素点データへの主成分・因子分析はその適用妥当性に欠けると認識されている(Linacre[9], 1998a, p. 603; Wright[10], 1996, p. 509)。データのRaschモデル適合度自体が、測定されたテスト(質問紙)構成概念についての一次元性充足度とみなされ、その潜在特性上での研究推察に関する一般化普遍性に具するデータ尺度構成の有用度を示すものとなる(Wright[5], 1991, p. 158、参照)。但し、データのRaschモデル適合度判断に際して一般的に利用されるRaschモデル測定出力適合度指標値は平越・井澤[11](2008)においても観察された様に完全無欠ではない。又、Raschモデル測定(以降、Rasch測定)の上では、『Raschモデル母数はデータが常にモデルへの“完全な”適合を果しているとの仮定に基づいて推定されている』(Linacre[12], 1999a, p. 676)ことにより、Rasch測定誤差としての標準化残差の小値無作為性がモデル適合を示すものとなる(Linacre, 1998b[13], p. 603 and 1999b[14], p. 710)。従って、Rasch測定においては、正誤反応データ自体ではなくRasch測定適用後における標準化残差の主成分分析が、特異項目機能(differential item functioning: DIF)察知を含めた包括的なデータ一次元性検証法として位置付けられている(Bond and Fox[15], 2001, pp. 155-156; Smith, Jr.[16], 2002; Tennant and Pallant[17], 2006; Wright and Stone[18], 2004, p. 21)。なお、ここでいう一次元性とは、分析対象の項目困難度と受験者能力の程度は同じRasch測定尺度で規定可能であり、その尺度は唯一の変数で表現できることを指している。

本稿では、一次元項目反応理論適用是非についての判断において、その前提とされる正誤反応データの主成分分析第一主成分寄与率(以降、単に、第一主成分寄与率)の下限値20% (Reckase[1], 1979, p. 227; 豊田[8]、2002、p. 30)はデータの一次元性充足度を観る絶対的な臨界値ではないことを例証する。第一主成分寄与率が項目困難度と受験者能力の分布によって大きく変動することを示し、前述のデータ一次元性判断基準を満たしていなくとも、現実テストのRasch測定によりデータ一次元性充足度が低いとは判定されない例が多数存在する可能性について考察する。先ず、一次元性を有しているテストにおける項目困難度と受験者能力の分布状況によって、第一主成分寄与率がどのように振舞うかについて述べる。次いで、現実に実施されるテストがどのような項目困難度と受験者能力の分布を持つのかという議論から、正誤反応データに関する一次元性充足度の判断基準値としてのその項目群第一主成分寄与率下限臨界値20%が妥当であるかを検討する。併せて、現実テストの一次元性充足度を確認する一方法について言及する。

## Ⅱ．項目反応理論一母数モデルにおける一次元性

### 1．項目反応理論一母数モデルの項目特性曲線

項目反応理論一母数モデルでは、項目困難度および受験者能力の二つの母数から導かれる正答確率を表すため、以下のような項目特性曲線を定義している。

$$P_j(\theta_i)=\frac{1}{1+e^{-(\theta_i-\beta_j)}}$$

$$\theta_i:\text{受験者}i\text{の能力値} \tag{1}$$

$$\beta_j:\text{項目}j\text{の困難度}$$

この項目特性曲線は、一母数ロジスティックモデル(渡辺・野口[7]、1999、p. 8)と呼ばれ、

$$P_j(\theta_i)=\frac{e^{\theta_i-\beta_j}}{e^{\theta_i-\beta_j}+e^{\theta_i-\beta_j}e^{-\theta_i+\beta_j}}=\frac{e^{\theta_i-\beta_j}}{e^{\theta_i-\beta_j}+1} \tag{2}$$

とあらわされる。Log オッズ比であるロジットコレクト（Logit Correct: 正答確率を誤答確率で除して対数をとったもの）は、

$$\ln\left(\frac{P_j(\theta_i)}{1-P_j(\theta_i)}\right)=\ln\left(\frac{e^{\theta_i-\beta_j}}{e^{\theta_i-\beta_j}+1}\frac{e^{\theta_i-\beta_j}+1}{1}\right)=\theta_i-\beta_j \tag{3}$$

となり、簡単に$\theta_i$、$\beta_j$の和差だけで表わされると示すことができる。なお、$\ln(\cdot)$は自然対数を表す。この定義によって、項目困難度と受験者能力を一次元の間隔尺度として表していることがわかる。

## 2. 項目反応理論一母数モデルの一次元性判断

受験者能力と項目困難度が間隔尺度として配置されるということは、受験者能力と項目困難度が一次元であり、かつその間隔に意味があるということである。つまり、項目反応理論一母数モデルでは以下の条件を前提としているのである。

「項目困難度及び受験者能力は一次元であり、困難度及び能力が定められた場合には、その大小と間隔に意味が見出せる。」

実際にテストを行い、項目困難度と受験者能力を推定する場合には、項目反応理論一母数モデルが条件とする一次元性充足が必要となるが、現実的にこの必要条件を満たしているかどうかを判定することが難しいことも事実である。現実的な方法として、まず、テストが項目反応理論一母数モデルに適合しているとの前提で以って推定を行い、その結果の解析から困難度及び能力の一次元性の尺度となるような指標を取り出し、テストの一次元妥当性を吟味することが多く行われている。もちろん、テスト前に項目群と受験者群を十分に吟味することが必須であることは言うまでもない。

実際のテスト運用者あるいは項目反応理論適用者が、あるテストの一次元性充足度において項目反応理論モデル適用が妥当な水準であるかどうかを確認する方法として、次のような基準が多く用いられている。

・各受験者と各項目を行と列に配置し、正誤を行列形式で記述した正誤反応行列から、各項目の相関係数行列を基に主成分分析を行い第一主成分寄与率が一定以上（多くは 20%前後以

上）であれば項目反応理論適用が妥当といえる程度の一次元性を有する。

第一主成分寄与率は、各項目間あるいは各受験者間で、正誤反応を仮に一次元の尺度に縮約した際に、その尺度で正誤反応のどの程度を説明することが可能であるかと考えることができるので、大きければ大きいほど一次元性が充足されていると考えることは間違いではない。しかしながら、判断の簡便さゆえに第一主成分寄与率の判断程度を定数として捉えることは無理があることを次に示すこととする。

## ３．一次元性を有する正誤反応から得られる第一主成分寄与率

項目困難度及び受験者能力の一次元性を本来的に有するテストを実施した理想的な状況を想定し、得られる正誤反応行列から第一主成分寄与率を導出した場合に、項目困難度及び受験者能力の分布によって第一主成分寄与率の程度がどのように変化するのかを考える。項目困難度と受験者能力は同様の議論がどちらにおいても成立する対称性を持っているため、ここでは項目困難度についての議論を行う。

本来の項目困難度の分布は無限に想定されるが、代表的なものとして、ここではまず各項目困難度が等間隔に一様配置され、分布幅の大小が違うという場合を考える。ただし、受験者の能力分布は一定として考える。項目困難度の分布幅が小さいということは、各項目間でほとんど困難度の差異がない状態である。このような場合には、図 **2-1** のように正答（1 で標示）と誤答（０で標示）が表現されることとなる。以降、正誤反応行列を０１データ行列、あるいは０１データと呼ぶこともある。項目困難度の分布範囲が大きくなるにつれ図 **2-2**、図 **2-3** のように正誤反応行列の様子は変化する。図 **2-1** から図 **2-3** において、「ほぼ全て０」あるいは「ほぼ全て１」の領域では、正答確率が非常に低い、あるいは非常に高い状態であり、概ね０か１の値をとるが、不規則に少数の逆の値が出現することがある状態である。「０、１がランダム」の領域では、正答確率が **0.5** に近く、０、１が同じ程度の生起確率を持っている状態であり、ランダムにほぼ同じ頻度で０と１が出現する状態である。図 **2-1** の $r_1$、図 **2-2** の $r_2$ の幅は、受験者能力分布幅が大きければ大きいほど小さくなる。また、図 **2-3** の $r_3$ の幅は、項目困難度分布幅が大きければ大きいほど小さくなる。ただし、ここでは受験者能力の中央値と項目困難度の中央値がほぼ同じ状態を想定しているが、違っていても今後の議論に大きな差は無い。

これらの正誤反応行列から各項目間の相関係数行列を求めると、図 **2-1** のように項目困難度分布幅が小さい場合には、全項目の正誤パターンは非常に近いものとなり、その結果相関係数行列はどの項目間においても大きな値をとることとなる。図 **2-4** に、項目困難度の分布幅が小さいときの項目間相関係数行列の概要を示す。図 **2-5**、図 **2-6** はそれぞれ、項目困難度分布幅が中庸であるとき、大きいときの項目間相関係数行列の概要である。図 **2-5** の楕円状の線は相関係数の等高線の概要を示している。これらの相関係数行列は、主成分分析において、次元の縮約に直接用いられるので、全般的に相関が高ければ第一主成分寄与率は高くなり、逆に全般的に相関が低ければ、各項目がそれぞれ独立していると考えられ、次元を縮約しても全体の構成を表すことが困難となるので第一主成分寄与率は低

くなる。以上により、項目困難度の分布範囲が小さいときには各項目間の相関が高くなり、そのために一次元に縮約しても第一主成分寄与率は大きくなるが、逆に項目困難度の分布範囲が大きいときには各項目間の相関は低くなり次元の縮約が困難であるが故に、一次元に縮約した際の第一主成分寄与率は小さくなることが明らかである。

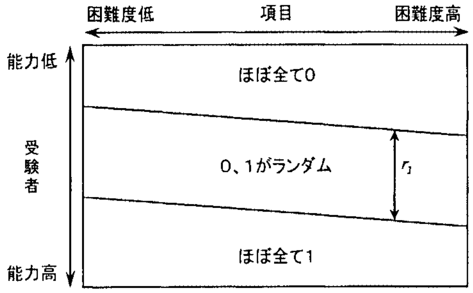


図 2-1　正誤反応行列（困難度範囲小）

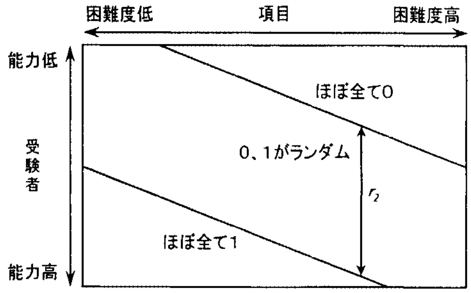


図 2-2　正誤反応行列（困難度範囲中）

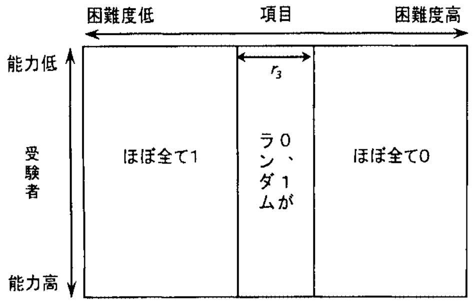


図 2-3　正誤反応行列（困難度範囲大）

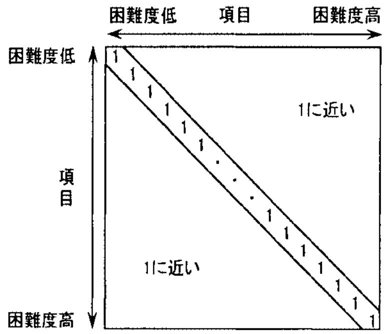


図 2-4　項目間相関行列（困難度範囲小）

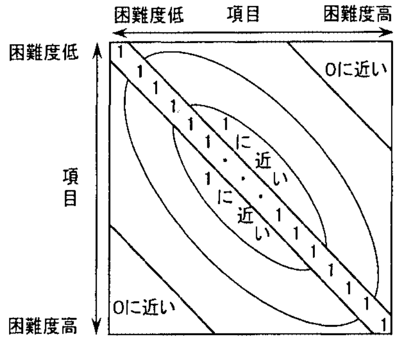


図 2-5　項目間相関行列（困難度範囲中）

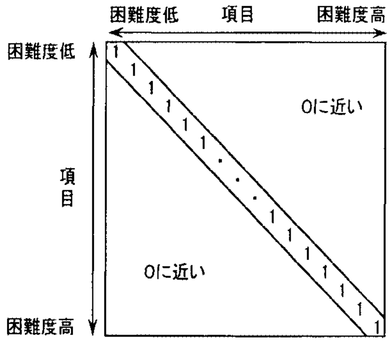


図 2-6　項目間相関行列（困難度範囲大）

## 4．第一主成分寄与率の例証

ここでは、一次元性を有するテストを実施したことを想定し、シミュレーションによって具体的な

第一主成分寄与率の変化を確認する。Ⅱ－3との関連性を考慮し、次のようなシミュレーションを行った。

- 項目数は 201、受験者数は 501 とする。
- 項目困難度と受験者能力の分布は等間隔で一様な分布とする。
- 項目困難度と受験者能力の分布範囲は、それぞれ(A) －1.5～1.5、(B) －3.0～3.0、(C)－6.0～6.0、(D) －12.0～12.0 とし、組合せ総数 16 通りのテストにおいて、正答確率を基にそれぞれの正誤反応を乱数を用いて 5 回ずつシミュレーションし、項目間相関係数行列から第一主成分寄与率を求め、その平均を算出する。
- 各テストにおいては、項目困難度と受験者能力が完全に一次元性を有しており、困難度と能力のみにおいて式(1)の正答確率が決定されることとする。

また、このシミュレーションとは別に、各項目・各受験者の正答確率から各項目間の期待相関係数を求めて算出した第一主成分寄与率の期待値も同時に求めた。なお、期待相関係数の導出方法については別途Ⅱ－6に記すこととする。シミュレーションによる第一主成分寄与率の平均値を表 2-1 に、第一主成分寄与率の期待値を表 2-2 に示す。なお、表 2-1 においては、シミュレーションデータによっては、全受験者が正答、あるいは誤答となる項目が発生することがあるが、その場合は該当項目を取り除いてから算出している。

表 2-1　第一主成分平均寄与率（シミュレーション）

| | | 項目困難度分布幅 | | | |
|---|---|---|---|---|---|
| | | -1.5～1.5 | -3～3 | -6～6 | -12～12 |
| 受験者能力分布幅 | -1.5～1.5 | 14.3% | 10.5% | 6.1% | 5.6% |
| | -3～3 | 36.8% | 29.9% | 17.9% | 14.1% |
| | -6～6 | 64.5% | 58.4% | 42.5% | 26.1% |
| | -12～12 | 82.0% | 78.3% | 68.5% | 47.8% |

表 2-2　第一主成分平均寄与率（期待値）

| | | 項目困難度分布幅 | | | |
|---|---|---|---|---|---|
| | | -1.5～1.5 | -3～3 | -6～6 | -12～12 |
| 受験者能力分布幅 | -1.5～1.5 | 13.9% | 10.4% | 6.0% | 3.2% |
| | -3～3 | 36.8% | 29.8% | 17.8% | 9.2% |
| | -6～6 | 64.5% | 58.3% | 42.5% | 22.5% |
| | -12～12 | 81.9% | 78.1% | 68.5% | 47.8% |

表 2-1、表 2-2 から、一次元性を本来的に有するテストにおいても、項目困難度分布幅、あるいは受験者能力分布幅が変われば、第一主成分寄与率は大きく変化することがわかる。この結果は、「どのようなテストにおいても第一主成分寄与率が一定以上でなければ項目反応理論による分析が妥当といえるだけの一次元性を有していない」との主張を否定する。つまり、現実的には項目反応理論による

分析が妥当といえるだけの一次元性を有しているかどうかは、第一主成分寄与率の状況による変化を勘案しながら、他の方法を用いて補完する必要性があると考えられる。

## 5．分布が正規分布の場合の第一主成分寄与率

これまでは項目困難度および受験者能力の分布については一様分布を仮定して議論を進めてきたが、実際のテストでは正規分布を仮定したほうが妥当性の高いことも多い。正規分布の場合でも、これまでの議論は有効であり、一様分布との違いを以下のように考えることで説明することができる。

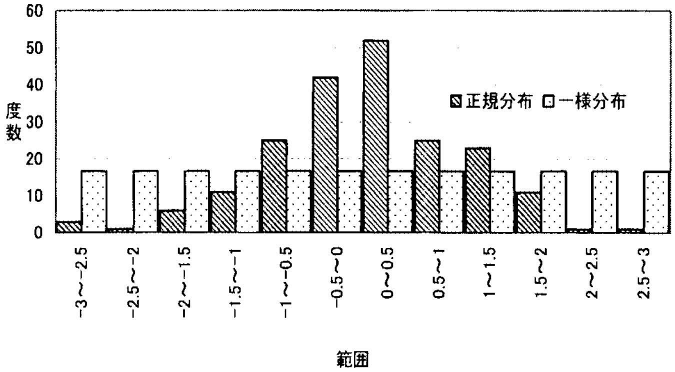


図2-7　一様分布と正規分布の項目数ヒストグラム例

正規分布は－∞～∞の変数範囲を持つので分布幅を有限とすることはできないが、例えば、標準偏差を用いたり、ある小さな確率密度以上をもつ範囲を想定したり、あるいは累積確率で範囲を想定したりすることが可能であるので、これらを用いて分布幅に準じるものを考えることができる。ここでは標準偏差を用いて項目困難度分布を考えることにする。項目困難度の分布が正規分布とみなせ、等間隔一様分布と同様の標準偏差を持つとすると、項目数 **201** の場合の一様分布と正規分布の困難度ヒストグラムは、例えば図 **2-7** のようになる。

一様分布と正規分布を比べると、正規分布では平均値付近の頻度が多く、分布の裾になればなるほど頻度が小さいことが明らかである。これは、同範囲の一様分布と比較した場合には、困難度分布の範囲が実質的に小さくなっていることと同じ影響を与えることがわかる。つまり、図 **2-7** で比較した場合、一様分布では各困難度範囲に均等に分布している設問のうち、両端の項目を中央に再配置するのと同様となり、正規分布では、より小さな範囲に設問が配置されている効果が出ることになる。結果的に困難度分布が同範囲の一様分布と比べた場合には、分布範囲のより狭い一様分布と同等の第一主成分寄与率が得られることになるのである。

次に具体的に、項目困難度分布が一様分布と同等の標準偏差を持つ場合における第一主成分寄与率を導出する。正規分布では、前述のとおり範囲が有限ではないので、各項目困難度をどのように設定するかという方法は様々に考えられる。ここでは正規乱数を発生させて各項目困難度に対応させ、正答確率行列からさらに乱数で正誤反応行列を生成し、これの第一主成分寄与率を算出する方法と、正

規分布をより忠実に再現できるよう、正規分布に沿うように各項目困難度を設定した上で、正答確率行列を基にした期待相関係数から第一主成分寄与率を算出する方法の、二つの方法を用いて検討することにする。

正規乱数によるシミュレーションは、次の通りであり、シミュレーション結果を表 2-3 に示す。

- 項目数は 201、受験者数は 501 とする。
- 項目困難度分布は平均 0、標準偏差を指定した正規乱数を用いる。
- 受験者能力は等間隔で一様な分布とする。
- 項目困難度の標準偏差は、(i) 0.870、(ii) 1.74 、(iii) 3.48、 (iv) 6.96 とする。これらはそれぞれⅡ－4に示した項目困難度分布(A)、(B)、(C)、(D)の標準偏差である。
- 受験者能力の分布範囲は、それぞれ(A) －1.5～1.5、(B) －3.0～3.0、(C) －6.0～6.0、(D) －12.0～12.0 とする。
- 困難度、能力の組合せ総数 16 通りのテストそれぞれで、各困難度決定において正規乱数を用いて 5 回ずつシミュレーションを行い、正答確率を基に乱数で生成したそれぞれの正誤反応行列から項目間相関係数を求め、さらに第一主成分寄与率を求めてその平均を算出する。
- 各テストにおいては、項目困難度と受験者能力が完全に一次元性を有しており、困難度と能力のみにおいて式(1)の正答確率が決定されることとする。
- 全ての受験者が正答する、あるいは誤答する項目がある場合にはその項目は取り除いている。

また、正規乱数の代わりに累積確率分布を等分するような方法で困難度を設定し、期待相関係数行列から第一主成分寄与率を求める方法では、前述のシミュレーションによる方法とつぎの点で異なる。

- 正規分布 N(0, $\sigma$ )を考える。これを事前設定正規分布と呼ぶことにする。
- 発生させたい項目困難度数 $N$ (ここでは 201 個)を用いて、全体を 1～ $N$ と番号付けし、この番号を項目困難度番号として、 $i$ と表現する。
- 各項目困難度を、次のように設定する。

　　困難度 $i = N^{-1}(\frac{i}{N+1}, 0, \sigma)$

- $\sigma$ を変化させて、設定される実際の困難度分布の標準偏差を目的の値にする。

ただし、 $N^{-1}(y, m, \sigma)$ は、事前設定正規分布の累積分布逆関数であり、 $y$ が累積確率を、 $m$ が平均値を、 $\sigma$ が事前設定正規分布標準偏差をそれぞれ表す。

この方法では、正規分布の累積分布の 0 と 1 の間を「発生させたい数+1」で等間隔に分割し、$-\infty$ および $\infty$ となる両端を除いて、それぞれに対応する値を使用することとなる。正規乱数のように現実的な振れやばらつきはないが、正規分布での傾向を調べるのが目的であるので、その意味においては正規乱数を用いるよりも安定した結果が期待できる。

次に、等間隔一様分布との標準偏差整合性を考え、事前設定正規分布標準偏差 $\sigma$ の設定から第一主成分寄与率の導出を、次のように行う。

・II－4 で用いた項目困難度の等間隔一様分布 4 種と同じ標準偏差を持つように、$\sigma$を 4 種類算出し、項目困難度分布を正規分布に入れ替えて、正答確率行列から第一主成分寄与率期待値を導出する。

まず、設定する正規分布項目困難度が、等間隔一様分布と同じ標準偏差をもつように$\sigma$を算出すると、表 2-4 のようになる。これをもとに、事前設定正規分布標準偏差$\sigma$を表 2-4 のように設定した上で、項目困難度分布にこれらの分布を適用した第一主成分寄与率を表 2-5 に示す。

表 2-3 困難度分布が正規分布の場合の第一主成分平均寄与率（シミュレーション）

| | | 項目困難度分布標準偏差（正規分布） | | | |
|---|---|---|---|---|---|
| | | 0.870 | 1.74 | 3.48 | 6.96 |
| 受験者能力分布幅 | -1.5～1.5 | 14.2% | 11.2% | 7.7% | 6.2% |
| | -3～3 | 36.9% | 31.2% | 18.4% | 16.3% |
| | -6～6 | 64.5% | 59.0% | 45.0% | 34.2% |
| | -12～12 | 81.9% | 78.6% | 64.0% | 55.2% |

表 2-4 項目困難度(201 個)の分布と標準偏差

| 等間隔一様分布 | | 正規分布 | |
|---|---|---|---|
| 設定 | 標準偏差 | 事前設定 | 標準偏差 |
| -1.5～1.5等間隔 | 0.87034 | N(0,0.89049) | 0.87034 |
| -3.0～3.0等間隔 | 1.74069 | N(0,1.78099) | 1.74069 |
| -6.0～6.0等間隔 | 3.48138 | N(0,3.56198) | 3.48138 |
| -12.0～12.0等間隔 | 6.96276 | N(0,7.12395) | 6.96276 |

表 2-5 困難度分布が正規分布の場合の第一主成分平均寄与率（期待値）

| | | 項目困難度分布標準偏差（正規分布） | | | |
|---|---|---|---|---|---|
| | | 0.870 | 1.74 | 3.48 | 6.96 |
| 受験者能力分布幅 | -1.5～1.5 | 14.0% | 11.0% | 7.1% | 4.1% |
| | -3～3 | 36.9% | 30.7% | 20.7% | 11.8% |
| | -6～6 | 64.5% | 58.6% | 45.1% | 27.8% |
| | -12～12 | 81.9% | 78.3% | 69.1% | 51.3% |

表 2-1～表 2-3、表 2-5 を比べると、全体的な傾向は全く同様であり、数値の比較においても、標準偏差が特に大きい場合に、正規分布のほうが第一主成分寄与率において若干大きい傾向は認められるが、概ね同等の値を示していることがわかる。つまり、正規分布においても標準偏差を基にした議論であれば、一様分布と傾向にはほとんど差異がなく、標準偏差を基にした議論は、一様分布と同様に成立することがわかる。なお、これ以外のいくつかの分布でも、ほぼ同様の傾向を示し、異なる傾向となる分布は認められなかった。以上本章での議論をまとめると、以下のようになる。

一次元性を有するテスト実施データにおいて、項目群についての主成分分析を行うと、

- 項目困難度分布、受験者能力分布の標準偏差によって、主成分分析の第一主成分寄与率は大きく変動する。
  - A) 項目困難度分布標準偏差が大きくなると項目群についての第一主成分寄与率は小さくなる。
  - B) 受験者能力分布標準偏差が大きくなると項目群についての第一主成分寄与率は大きくなる。
- 第一主成分寄与率が小さい場合にも項目群についての一次元性が認められる場合がある。そのため第一主成分寄与率以外の考察が不可欠となる。
- 項目困難度及び受験者能力の分布が一様分布でなく正規分布である場合には同程度の標準偏差を持つ一様分布とほぼ同じ第一主成分寄与率となる。

なお、ここでは項目群についての主成分分析に対しての議論を行っているが、受験者群についても対称的に同様の議論が成立することは明らかである。また、第一主成分寄与率が低い場合においても、テストの一次元性が項目反応理論で分析を行う妥当性が備わっているかどうかの判断方法として、項目群と受験者群の分割による推定値不変性の調査等が挙げられる。これらは一次元性の確認方法とともに次章以降で述べることとする。

## 6．期待相関係数の導出

ここでは、表 2-2 及び表 2-5 に関連した、期待相関係数の導出方法を記す。

項目 $j_1$ と項目 $j_2$ の困難度をそれぞれ $\beta_{j1}$、$\beta_{j2}$ とする。また、受験者数を $n$ とし、各受験者の能力を $\theta_i (i=1,2,\cdots,n)$ とする。受験者 $i$ が項目 $j_1$ と項目 $j_2$ の両方に正答する確率は $P_{j1}(\theta_i)P_{j2}(\theta_i)$ であり、同様に項目 $j_1$ に正答、項目 $j_2$ に誤答する確率は $P_{j1}(\theta_i)(1-P_{j2}(\theta_i))$、項目 $j_1$ に誤答、項目 $j_2$ に正答する確率は $(1-P_{j1}(\theta_i))P_{j2}(\theta_i)$、項目 $j_1$ にも項目 $j_2$ にも誤答する確率は $(1-P_{j1}(\theta_i))(1-P_{j2}(\theta_i))$ となる。ただし、$P_j(\theta_i)$ は式(1)に示した $P_j(\theta_i)=\dfrac{1}{1+e^{-(\theta_i-\beta_j)}}$ である。これらより、項目 $j_1$ と項目 $j_2$ における分割表の期待値は表 2-6 のように計算することができる。

表 2-6　項目 $j_1$ と項目 $j_2$ の正誤頻度の分割表期待値

| | | 項目j1 | |
|---|---|---|---|
| | | 誤答 | 正答 |
| 項目j2 | 誤答 | $c00=\sum_{i=1}^{n}(1-P_{j1}(\theta_i))(1-P_{j2}(\theta_i))$ | $c10=\sum_{i=1}^{n}P_{j1}(\theta_i)(1-P_{j2}(\theta_i))$ |
| | 正答 | $c01=\sum_{i=1}^{n}(1-P_{j1}(\theta_i))P_{j2}(\theta_i)$ | $c11=\sum_{i=1}^{n}P_{j1}(\theta_i)P_{j2}(\theta_i)$ |

表 2-6 の分割表を基にすると、項目 $j_1$ と項目 $j_2$ の相関係数 $cc$ は、

$$cc=\frac{S_{j1j2}}{\sqrt{S_{j1j1}S_{j2j2}}} \tag{4}$$

となる。ただし、正答を1、誤答を0と考えて、項目ごとに分布を平均0、標準偏差1に標準化して、次の通り導出する。

$$\overline{C}_{j1}=(c00+c01)\times 0+(c10+c11)\times 1,\quad \overline{C}_{j2}=(c00+c10)\times 0+(c01+c11)\times 1,$$

$$V_{j1}=\frac{(0-\overline{C}_{j1})^2(c00+c01)+(1-\overline{C}_{j1})^2(c10+c11)}{n},$$

$$V_{j2}=\frac{(0-\overline{C}_{j2})^2(c00+c10)+(1-\overline{C}_{j2})^2(c01+c11)}{n},$$

$$N0_{j1}=\frac{(0-\overline{C}_{j1})}{V_{j1}},\quad N1_{j1}=\frac{(1-\overline{C}_{j1})}{V_{j1}},\quad N0_{j2}=\frac{(0-\overline{C}_{j2})}{V_{j2}},\quad N1_{j2}=\frac{(1-\overline{C}_{j2})}{V_{j2}},$$

$$S_{j1j1}=(c00+c01)N0_{j1}^2+(c10+c11)N1_{j1}^2,\quad S_{j2j2}=(c00+c10)N0_{j2}^2+(c01+c11)N1_{j2}^2,$$

$$S_{j1j2}=c00N0_{j1}N0_{j2}+c10N1_{j1}N0_{j2}+c01N0_{j1}N1_{j2}+c11N1_{j1}N1_{j2}. \tag{5}$$

このようにして、全ての項目間の期待相関係数を算出して期待相関係数行列を構成してから、主成分分析を行って第一主成分寄与率を求めたものが、表 **2-2** 及び表 **2-5** の第一主成分平均寄与率（期待値）である。

## Ⅲ. Rasch 母数推定値標準偏差の大きさによる第一主成分寄与率変動の例示

項目困難度と受験者能力の相互連関的な分布特徴により顕現される第一主成分寄与率の変動傾向一端を本章で明示する。前章での考察内容についての例示を一連の模擬データ群と現実テストによって順次に与える。その意図は、人間科学研究上でのデータ一次元性に関係する事象生起傾向の認知重要性に基づくものである。

### 1. 模擬データ群

項目群困難度と受験者群能力を表 **3-1** のように設定し、三種類の模擬データ群を作成する。表 **3-1** での模擬データ群作成要領で以って、ＡＢＣのそれぞれにおいて **1) 30** 項目対 **500** 名、**2) 60** 項目対 **500** 名、**3) 90** 項目対 **500** 名、**4) 120** 項目対 **500** 名、**5) 200** 項目対 **500** 名、**6) 250** 項目対 **500** 名を配置する。なお、ＡＢＣ各々における受験者群 **500** 名は同一設定値分布とされている。ＡとＢでの正規分布設定値出力は **MINITAB (Release 14, Minitab Inc., 2003)** によるものであり、その出力両端値を表 **3-1** に含めている。各項目と各受験者の設定値一対すべてに対する **Rasch** 項目分析モデル期待正答確率（前章第一節掲示式 **(2)**）に基づく出力値への **Excel** 上での **RAND** 関数適用で以って、ＡＢＣそれぞれにおける上記六つの正誤反応データが与えられる。その0と1の確率的な生起の上で、各データは完全な一次元性を有するものとなっている。表 **3-1** での作成要領に基づくＡＢＣそれぞれにおける各データに関する指標値一覧を下に与える。表 **3-2** から表 **3-4** の理解の上で、指標についての説明を

前もって以下に記述する。

表 3-1 三種類の模擬データ群作成要領

A) 項目群困難度設定値： (等間隔)一様分布

| | | 30項目 | 60項目 | 90項目 | 120項目 | 200項目 | 230項目 |
|---|---|---|---|---|---|---|---|
| 項目群事前設定値 | 最小値 | -4.000 | -4.000 | -4.000 | -4.000 | -4.000 | -4.000 |
| | 刻み目 | 0.270 | 0.135 | 0.090 | 0.068 | 0.040 | 0.032 |
| | 最大値 | 3.830 | 3.965 | 4.010 | 4.092 | 3.960 | 3.968 |

受験者群能力設定値： 標準正規分布 N (0, 1)

| | | |
|---|---|---|
| 受験者群事前設定値 | 最小値 | -2.884 |
| | 標準正規分布 | N (0, 1) |
| | 最大値 | 3.000 |

B) 項目群困難度設定値： 正規分布 N (0, 1.2)

| | | 30項目 | 60項目 | 90項目 | 120項目 | 200項目 | 250項目 |
|---|---|---|---|---|---|---|---|
| 項目群事前設定値 | 最小値 | -2.677 | -2.576 | -3.043 | -3.234 | -3.213 | -3.319 |
| | 正規分布 | N (0, 1.2) | N (0, 1.2) | N (0, 1.2) | N (0, 1.2) | N (0, 1.2) | N (0, 1.2) |
| | 最大値 | 2.821 | 2.796 | 2.680 | 3.168 | 3.301 | 3.361 |

受験者群能力設定値： 標準正規分布 N (0, 1)

| | | |
|---|---|---|
| 受験者群事前設定値 | 最小値 | -2.884 |
| | 標準正規分布 | N (0, 1) |
| | 最大値 | 3.000 |

C) 項目群困難度設定値： (等間隔)一様分布

| | | 30項目 | 60項目 | 90項目 | 120項目 | 200項目 | 250項目 |
|---|---|---|---|---|---|---|---|
| 項目群事前設定値 | 最小値 | -4.000 | -4.000 | -4.000 | -4.000 | -4.000 | -4.000 |
| | 刻み目 | 0.270 | 0.135 | 0.090 | 0.068 | 0.040 | 0.032 |
| | 最大値 | 3.830 | 3.965 | 4.010 | 4.092 | 3.960 | 3.968 |

受験者群能力設定値： (等間隔)一様分布

| | | |
|---|---|---|
| 受験者群事前設定値 | 最小値 | -3.000 |
| | 刻み目 | 0.012 |
| | 最大値 | 2.988 |

各表における第一主成分寄与率は、各データの項目群と受験者群についての SPSS (Version 15.0, SPSS Inc., 2006)によるピアソン相関係数に基づく主成分分析出力値である。線形関係にはない 0 1 データ列間へのピアソン相関係数出力はその適用に妥当性が欠けるとされ、応用理論の上では 0 1 データの主成分・因子分析はテトラコリック(四分)相関係数に基づく方法が望ましいと渡辺・野口[7](1999、p. 24)によって記されている。表 3-2～表 3-4 には MicroFACT 2.0 (Niels G. Waller, 2000, Psychometrics Software, distributed exclusively by Assessment System Corporation)により出力されたテトラコリック相関係数に基づく第一主成分寄与率を表の大きさ制限により含めてはいない。唯、正誤反応データへの主成分分析適用の上では、後者が前者を第一主成分寄与率として一般的に 5～20%の範囲内で常に上回り、データ一次元性充足度の下限信頼性の観点からは、ピアソン相関係数に

基づく主成分分析出力値参照で事が足りると思われる。

各データの項目困難度と受験者能力の Rasch 測定は Facets (Linacre, 1989-2001, Winsteps)により出力された推定値であり、項目群困難度と受験者群能力の推定値分布情報としての最小値、最大値、標準偏差、推定値範囲を表に含めている。項目群と受験者群に対して Rasch 測定上で Facets により出力されるそれぞれの分離信頼性係数は、Andrich[19] (1982)に由来しており、古典的テスト理論においてと同様に、[信頼性係数 = 真の得点分散 / 観測得点分散]との定義に基づくものである。その指標概念としては、項目群分離信頼性係数は受験者群によって困難度推定値が如何なる程度(0.00～1.00)に分離されているかというものであり、そのテスト受験者群母集団における当該困難度推定値の信頼性・再現性・同一順序性を示すものとなり、受験者群分離信頼性係数はその対称指標である(Fox and Jones[20], 1998, pp. 35- 37、参照)。能力推定値と困難度推定値のいずれについても、それぞれデータにおける項目数と受験者数の大きさに依存して推定値標準誤差が押し並べて小さな値になるにつれて、その分離信頼性係数は 1 .00 に近づくことがその数理上の特徴である。又、受験者群分離信頼性係数は、古典的テスト理論における Cronbach のアルファ係数あるいはそれと正誤反応データに関して同等な Kuder-Richardson Formula 20 で以って与えられる指標値とほぼ同値となる(Andrich[19] , 1982)。従って、表 3-2 における受験者群能力推定値についての分離信頼性係数が 01 データ上での項目群内的一貫性係数として SPSS によって出力される Cronbach のアルファ係数の値とみなされても良いことになる。受験者群分離信頼性係数と同様に、アルファ係数も項目数増大に大きく依存する指標である(Zhu, Updyke, and Lewandowski[21], 1997, p. 287)ことに留意される。受験者群と項目群の分離信頼性係数算出式は以下の通りである(Andrich[19] , 1982, p. 95; Linacre[22], 1989- 2001, p. 77; Wright and Masters[23], 1982, pp. 105- 106、参照)。

・受験者群分離信頼性係数 = 誤差調整済み能力推定値分散 / 能力推定値分散

= (能力推定値分散 - 受験者平均能力推定値誤差分散) / 能力推定値分散

= (能力推定値の分散 - 受験者群モデル標準誤差二乗和の平均値) / 能力推定値の分散

・項目群分離信頼性係数 = 誤差調整済み困難度推定値分散 / 困難度推定値分散

= (困難度推定値分散 - 項目平均困難度推定値誤差分散) / 困難度推定値分散

= (困難度推定値の分散 - 項目群モデル標準誤差二乗和の平均値) / 困難度推定値の分散

表 3-2 における項目群と受験者群に関する推定値不変性成立度は、平越・井澤 [11](2008)における表 4(p. 78)での提示と同じく、各データの受験者群と項目群それぞれについての折半法を受験者群と項目群それぞれに関する Rasch 測定標準化残差主成分分析負荷量記号正負分別に基づくものとしている。なお、既述の如く、当該模擬データ群作成要領により、各データは完全な一次元性を有するものとなっている。受験者群折半データ間同一項目群、及び、項目群折半データ間同一受験者群に対してそれぞれ Rasch 測定により与えられた困難度推定値と能力推定値に関するピアソン相関係数とケンドール順位相関係数タウｂが、項目群と受験者群それぞれについての推定値不変性成立度である。従って、

その相関係数は、各データにおける項目数と項目群困難度推定値分布、並びに、受験者数と受験者群能力推定値分布に関係する各データ母数推定値精度上での Rasch 測定妥当性・信頼性の高さを示す指標と理解される。

表 3-2 Aデータ群指標値一覧

| データ | 事前設定値 | 最小値 | 刻み目 | 最大値 | 01データ項目群ピアソン主成分分析第1主成分寄与率 | 01データ受験者群ピアソン主成分分析第1主成分寄与率 | Rasch測定Facets出力 | 最小値 | 最大値 | 標準偏差 | 範囲 | 推定値分離信頼性係数 | 推定値不変性 有効数 | ピアソン | ケンドールタウb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| データ1 (30 項目 x 500 名) | 項目群事前設定値 | -4.000 | 0.270 | 3.830 | 13.4% | 51.0% | 項目困難度推定値 | -4.22 | 4.13 | 2.57 | 8.35 | 1.00 | 30 | 0.929 | 0.791 |
| | 受験者群事前設定値 | -2.884 | 標準正規分布 N (0, 1) | 3.000 | | | 受験者能力推定値 | -4.02 | 3.60 | 1.22 | 7.62 | 0.79 | 500 | 0.581 | 0.468 |
| データ2 (60 項目 x 500 名) | 項目群事前設定値 | -4.000 | 0.135 | 3.965 | 11.7% | 48.6% | 項目困難度推定値 | -3.95 | 4.22 | 2.38 | 8.17 | 1.00 | 60 | 0.945 | 0.823 |
| | 受験者群事前設定値 | -2.884 | 標準正規分布 N (0, 1) | 3.000 | | | 受験者能力推定値 | -2.83 | 3.20 | 1.08 | 6.03 | 0.88 | 500 | 0.703 | 0.541 |
| データ3 (90 項目 x 500 名) | 項目群事前設定値 | -4.000 | 0.090 | 4.010 | 10.6% | 49.6% | 項目困難度推定値 | -3.94 | 4.23 | 2.40 | 8.17 | 1.00 | 90 | 0.955 | 0.838 |
| | 受験者群事前設定値 | -2.884 | 標準正規分布 N (0, 1) | 3.000 | | | 受験者能力推定値 | -3.86 | 3.05 | 1.03 | 6.91 | 0.91 | 500 | 0.744 | 0.571 |
| データ4 (120 項目 x 500 名) | 項目群事前設定値 | -4.000 | 0.068 | 4.092 | 10.8% | 49.6% | 項目困難度推定値 | -4.23 | 4.18 | 2.41 | 8.41 | 1.00 | 120 | 0.961 | 0.845 |
| | 受験者群事前設定値 | -2.884 | 標準正規分布 N (0, 1) | 3.000 | | | 受験者能力推定値 | -2.96 | 3.16 | 1.05 | 6.12 | 0.93 | 500 | 0.799 | 0.615 |
| データ5 (200 項目 x 500 名) | 項目群事前設定値 | -4.000 | 0.040 | 3.960 | 10.7% | 48.0% | 項目困難度推定値 | -4.60 | 3.97 | 2.33 | 8.57 | 1.00 | 200 | 0.966 | 0.855 |
| | 受験者群事前設定値 | -2.884 | 標準正規分布 N (0, 1) | 3.000 | | | 受験者能力推定値 | -3.34 | 3.06 | 1.03 | 6.40 | 0.96 | 500 | 0.855 | 0.669 |
| データ6 (250 項目 x 500 名) | 項目群事前設定値 | -4.000 | 0.032 | 3.968 | 10.6% | 47.8% | 項目困難度推定値 | -4.19 | 4.26 | 2.33 | 8.45 | 1.00 | 250 | 0.969 | 0.862 |
| | 受験者群事前設定値 | -2.884 | 標準正規分布 N (0, 1) | 3.000 | | | 受験者能力推定値 | -2.92 | 2.63 | 1.02 | 5.55 | 0.97 | 500 | 0.876 | 0.692 |

| | | | | | |
|---|---|---|---|---|---|
| 上掲6データ指標値範囲 | 10.6% ｜ 13.4% | 47.8% ｜ 51.0% | 項目困難度推定値不変性成立程度 | ピ ア ソ ン | 0.929 ～ 0.969 |
| | | | | ケンドールタウb | 0.791 ～ 0.862 |
| | | | 受験者能力推定値不変性成立程度 | ピ ア ソ ン | 0.581 ～ 0.876 |
| | | | | ケンドールタウb | 0.468 ～ 0.692 |

Aの 6 データ指標値一覧について顕著に観察されることは次の三点である。

1) 困難度推定値分布標準偏差がほぼ 2.40 前後の大きな値であり、前章第 5 節での論証との連関性で以って受験者群第一主成分寄与率はほぼ 49%前後のかなり高い値である。それに相応して、各データ項目群困難度推定値について大変に高い Rasch 測定妥当性・信頼性が示されている。

2) 1.07 前後の能力推定値標準偏差は大きな値ではなく中庸であり、項目群第一主成分寄与率はほぼ 11%前後の低い値である。それに呼応して、能力推定値についての Rasch 測定妥当性・信頼性はかなり高いとはみなされず、項目数が減少するにつれてその確かな低下傾向が明示されている。

3) 項目群分離信頼性係数(受験者群能力推定値再現性指標)は各データにおいて 1.00 であり、受験者群分離信頼性係数(項目群困難度推定値再現性指標)は項目数増加につれて 0.79 から 0.96 に渡って上昇している。

Bの 6 データ指標値一覧についてAでの各対応指標値と較べて顕著に観察されることは次の三点である。

表 3-3　　Bデータ群指標値一覧

| データ | 群 | 事前設定値 最小値 | 事前設定値 分布 | 事前設定値 最大値 | 01データ項目群ピアソン主成分分析第1固有値寄与率 | 01データ受験者群ピアソン主成分分析第1固有値寄与率 | Rasch測定Facets出力 | 最小値 | 最大値 | 標準偏差 | 範囲 | 推定値分離信頼性係数 | 推定値不変性 有効数 | ピアソン | ケンドールタウb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| データ1 (30項目 x 500名) | 項目群 | -2.677 | 正規分布 N(0, 1.2) | 2.821 | 17.6% | 24.6% | 項目困難度推定値 | -2.96 | 2.72 | 1.33 | 5.68 | 0.99 | 30 | 0.807 | 0.609 |
| | 受験者群 | -2.884 | 標準正規分布 N(0, 1) | 3.000 | | | 受験者能力推定値 | -4.13 | 4.07 | 1.17 | 8.20 | 0.84 | 493 | 0.614 | 0.475 |
| データ2 (60項目 x 500名) | 項目群 | -2.576 | 正規分布 N(0, 1.2) | 2.796 | 16.5% | 19.3% | 項目困難度推定値 | -2.67 | 2.82 | 1.17 | 5.49 | 0.99 | 60 | 0.840 | 0.639 |
| | 受験者群 | -2.884 | 標準正規分布 N(0, 1) | 3.000 | | | 受験者能力推定値 | -2.89 | 4.00 | 1.07 | 6.89 | 0.91 | 498 | 0.753 | 0.589 |
| データ3 (90項目 x 500名) | 項目群 | -3.043 | 正規分布 N(0, 1.2) | 2.680 | 15.4% | 22.3% | 項目困難度推定値 | -3.17 | 2.87 | 1.26 | 6.04 | 0.99 | 90 | 0.883 | 0.737 |
| | 受験者群 | -2.884 | 標準正規分布 N(0, 1) | 3.000 | | | 受験者能力推定値 | -2.62 | 3.07 | 1.04 | 5.69 | 0.94 | 500 | 0.810 | 0.640 |
| データ4 (120項目 x 500名) | 項目群 | -3.234 | 正規分布 N(0, 1.2) | 3.168 | 14.9% | 23.2% | 項目困難度推定値 | -3.25 | 3.16 | 1.32 | 6.41 | 0.99 | 120 | 0.921 | 0.765 |
| | 受験者群 | -2.884 | 標準正規分布 N(0, 1) | 3.000 | | | 受験者能力推定値 | -2.80 | 3.05 | 1.02 | 5.85 | 0.95 | 500 | 0.841 | 0.668 |
| データ5 (200項目 x 500名) | 項目群 | -3.213 | 正規分布 N(0, 1.2) | 3.301 | 14.6% | 20.5% | 項目困難度推定値 | -3.43 | 3.75 | 1.22 | 7.18 | 0.99 | 200 | 0.925 | 0.777 |
| | 受験者群 | -2.884 | 標準正規分布 N(0, 1) | 3.000 | | | 受験者能力推定値 | -3.26 | 2.78 | 0.99 | 6.04 | 0.97 | 500 | 0.888 | 0.707 |
| データ6 (250項目 x 500名) | 項目群 | -3.319 | 正規分布 N(0, 1.2) | 3.361 | 14.8% | 20.8% | 項目困難度推定値 | -3.38 | 3.32 | 1.22 | 6.70 | 0.99 | 250 | 0.937 | 0.786 |
| | 受験者群 | -2.884 | 標準正規分布 N(0, 1) | 3.000 | | | 受験者能力推定値 | -3.14 | 2.72 | 1.00 | 5.86 | 0.98 | 500 | 0.902 | 0.733 |

| | | | | | |
|---|---|---|---|---|---|
| 上掲6データ指標値範囲 | 14.6% ｜ 17.6% | 19.3% ｜ 24.6% | 項目困難度推定値不変性成立程度 | ピアソン | 0.807 ～ 0.937 |
| | | | | ケンドールタウb | 0.609 ～ 0.786 |
| | | | 受験者能力推定値不変性成立程度 | ピアソン | 0.614 ～ 0.902 |
| | | | | ケンドールタウb | 0.475 ～ 0.733 |

1) Bでの1.05前後の能力推定値標準偏差とAでのその対応値1.07はほぼ同等であるが、Bでの1.25前後の困難度推定値標準偏差はAでのその対応値2.40のほぼ半分である。それに準じて、Bでの22%前後の受験者群第一主成分寄与率が、Aでのその対応値49%に較べて半分以下となっている。それに呼応して、Bにおける困難度推定値についてのRasch測定妥当性・信頼性がAにおけるそれよりも全般的に下回っている。

2) 前章第5節での論証により、能力推定値標準偏差が同等であるならば、項目群困難度推定値標準偏差が小さくなれば項目群第一主成分寄与率は大きくなる。それで以って、Bでの項目群第一主成分寄与率は、20%未満ながらも、Aでのその対応値より4%ばかり上回っている。それに呼応して、Bでの能力推定値についてのRasch測定妥当性・信頼性はAのそれよりも全般的に僅かながらに上回っている。

3) Bでの項目群分離信頼性係数(受験者群能力推定値再現性指標)はAでの対応値より0.01下回って各データに0.99が付されて、Bでの受験者群分離信頼性係数(項目群困難度推定値再現性指標)はAでの各データ対応値を上回って0.84から0.97となっている。

Cの6データ指標値一覧についてAでの各対応指標値と較べて顕著に観察されることは次の三点である。

表 3-4 Cデータ群指標値一覧

| データ | 事前設定値 | 最小値 | 刻み目 | 最大値 | 01データ項目群ピアソン主成分分析第1固有値寄与率 | 01データ受験者群ピアソン主成分分析第1固有値寄与率 | Rasch測定Facets出力 | 最小値 | 最大値 | 標準偏差 | 範囲 | 推定値分離信頼性係数 | 有効数 | 推定値不変性 ピアソン | 推定値不変性 ケンドールタウb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| データ1 (30項目 x 500名) | 項目群事前設定値 | -4.000 | 0.270 | 3.830 | 27.0% | 43.8% | 項目困難度推定値 | -4.25 | 4.23 | 2.55 | 8.48 | 1.00 | 30 | 0.919 | 0.788 |
| | 受験者群事前設定値 | -3.000 | 0.012 | 2.988 | | | 受験者能力推定値 | -4.59 | 4.57 | 1.95 | 9.16 | 0.91 | 491 | 0.767 | 0.626 |
| データ2 (60項目 x 500名) | 項目群事前設定値 | -4.000 | 0.135 | 3.965 | 25.6% | 42.6% | 項目困難度推定値 | -4.00 | 4.27 | 2.44 | 8.27 | 1.00 | 60 | 0.949 | 0.826 |
| | 受験者群事前設定値 | -3.000 | 0.012 | 2.988 | | | 受験者能力推定値 | -4.16 | 3.69 | 1.82 | 7.85 | 0.95 | 498 | 0.866 | 0.711 |
| データ3 (90項目 x 500名) | 項目群事前設定値 | -4.000 | 0.090 | 4.010 | 24.7% | 42.7% | 項目困難度推定値 | -4.20 | 3.85 | 2.40 | 8.05 | 1.00 | 90 | 0.958 | 0.826 |
| | 受験者群事前設定値 | -3.000 | 0.012 | 2.988 | | | 受験者能力推定値 | -3.73 | 3.72 | 1.77 | 7.45 | 0.97 | 500 | 0.905 | 0.752 |
| データ4 (120項目 x 500名) | 項目群事前設定値 | -4.000 | 0.068 | 4.092 | 24.1% | 42.4% | 項目困難度推定値 | -4.09 | 4.19 | 2.36 | 8.28 | 1.00 | 120 | 0.967 | 0.846 |
| | 受験者群事前設定値 | -3.000 | 0.012 | 2.988 | | | 受験者能力推定値 | -3.50 | 3.60 | 1.72 | 7.10 | 0.97 | 500 | 0.919 | 0.774 |
| データ5 (200項目 x 500名) | 項目群事前設定値 | -4.000 | 0.040 | 3.960 | 25.2% | 41.8% | 項目困難度推定値 | -4.18 | 4.21 | 2.35 | 8.39 | 1.00 | 200 | 0.973 | 0.861 |
| | 受験者群事前設定値 | -3.000 | 0.012 | 2.988 | | | 受験者能力推定値 | -3.56 | 3.38 | 1.78 | 6.94 | 0.99 | 500 | 0.944 | 0.806 |
| データ6 (250項目 x 500名) | 項目群事前設定値 | -4.000 | 0.032 | 3.968 | 24.7% | 41.7% | 項目困難度推定値 | -4.23 | 4.51 | 2.33 | 8.74 | 1.00 | 250 | 0.975 | 0.865 |
| | 受験者群事前設定値 | -3.000 | 0.012 | 2.988 | | | 受験者能力推定値 | -3.24 | 3.34 | 1.74 | 6.58 | 0.99 | 500 | 0.955 | 0.826 |

| 上掲6データ指標値範囲 | 項目群寄与率 | 受験者群寄与率 | | | |
|---|---|---|---|---|---|
| | 24.1% ｜ 27.0% | 41.7% ｜ 43.8% | 項目困難度推定値不変性成立程度 | ピアソン | 0.919 ～ 0.975 |
| | | | | ケンドールタウb | 0.788 ～ 0.865 |
| | | | 受験者能力推定値不変性成立程度 | ピアソン | 0.767 ～ 0.955 |
| | | | | ケンドールタウb | 0.626 ～ 0.826 |

1) Aでの各データと同じ項目群を持つCでの困難度推定値標準偏差はAと同じくほぼ 2.40 前後であるが、Cでのほぼ 1.80 前後の能力推定値標準偏差はAでのその対応値 1.07 よりかなり大きい値である。前章第 5 節での論証が例示されて、Cでの項目群第一主成分寄与率としての 25%前後の値はAでのその対応値 11%の二倍強である。それに準じて、Aに比べてCでの能力推定値についての Rasch 測定妥当性・信頼性が全般的に 14%程度高くなっている。

2) Aでの各データと同じ項目群を持つCでの困難度推定値標準偏差はAと同じくほぼ同等に 2.40 前後であるから、能力推定値標準偏差が大きくなれば受験者群第一主成分寄与率は小さくなる。その例示として、Cでのほぼ 42.5%の受験者群第一主成分寄与率はAでのその対応値 49%を下回っている。しかし、その差があるとしても両者の受験者群第一主成分寄与率自体がかなり高いことで以って、Cでの困難度推定値についての Rasch 測定妥当性・信頼性はAとほぼ同程度に大変に高いと示されており、それについてはCとAの間に差はない。

3) Cでの項目群分離信頼性係数(受験者群能力推定値再現性指標)はAでの対応値 1.00 と同じく各データに 1.00 が付されて、C での受験者群分離信頼性係数(項目群困難度推定値再現性指標)はAでの各データ対応値をかなりの程度に上回って 0.91 から 0.99 となっている。

ABCでの指標値比較まとめとして、次の三点が明確に観察される。

1) 項目群第一主成分寄与率についてはＣでの25%前後の値がＡとＢでの20%未満の値に優っている。その主たる要因は、前章第5節で論証された能力推定値標準偏差の大きさであり、Ｃの全般的にＡとＢに優る能力推定値標準偏差との連関性である。受験者群第一主成分寄与率についてはＡとＣが圧倒的にＢを上回っており、ＡとＣでのその値はそれぞれほぼ49%、42.5%である。ＡとＣでの困難度推定値標準偏差はほぼ同等に大きな値として2.40前後であり、前章第5節での受験者群第一主成分寄与率に関する大きな困難度推定値標準偏差と大きな受験者群第一主成分寄与率との連関性が認められる。一方、ＡとＣでの困難度推定値標準偏差はそれぞれほぼ1.07、1.80であり、能力推定値標準偏差が大きくなると受験者群第一主成分寄与率は小さくなるという前章第5節での論証が例示されている。

2) 困難度推定値についてのRasch測定妥当性・信頼性については、ＡとＣがＢに優ってほぼ同等に大変に高い程度と示されており、ＡとＣでのＢにはるかに優る受験者群第一主成分寄与率の大きさに呼応している。従って、困難度推定値標準偏差が大きくなれば、受験者群第一主成分寄与率は大きくなるとの前章第5節での論証との連関性で以って、困難度推定値標準偏差が大きくなれば、困難度推定値についてのRasch測定妥当性・信頼性が高くなる傾向一端が窺われる。又、ＡＢＣにおいて僅かにしても項目数の増大が困難度推定値についてのRasch測定妥当性・信頼性を高めていることが観察される。能力推定値についてのRasch測定妥当性・信頼性については、ＡＢＣの順序でそれが高くなっており、その順序は項目群第一主成分寄与率の大きさに呼応している。能力推定値標準偏差が大きくなれば、項目群第一主成分寄与率は大きくなるとの前章第5節での論証との連関性で以って、能力推定値標準偏差が大きくなれば、能力推定値についてのRasch測定妥当性・信頼性が高くなる傾向一端が窺われる。又、能力推定値についてのRasch測定妥当性・信頼性は項目数の増加によって顕著に高められることがＡＢＣによって顕示されている。

3) Rasch測定分離信頼性係数については、項目群と受験者群に関するそれぞれが困難度推定値と能力推定値についてのRasch測定妥当性・信頼性に呼応することがＡＢＣにおいて認められる。然しながら、後者に較べて前者の変動域がはるかに狭いと示されている。これは、Rasch測定分離信頼性係数のRasch測定母数推定値分布異同に対する検知鋭敏性欠如傾向の示唆である。

ちなみに、前掲表3-2から表3-4における受験者群500名に関係するデータ群に、前掲表3-1でのＡＢＣ分布設定値の上で受験者数100名、250名、1,000名から成るデータ群を加えて、ＡＢＣにおける項目群と受験者群それぞれについてのピアソン相関係数に基づく第一主成分寄与率の変動範囲を与えるものが表3-6である。参考までに、ＡＢＣそれぞれにおける項目群についてのCronbachのアルファ係数、及び、テトラコリック相関係数に基づく主成分分析の上での第一主成分寄与率についての変動範囲もその表に含めている。前者と後者はそれぞれSPSSとMicroFACTによって出力された値に基づいている。なお、ＡＢＣそれぞれにおけるデータ群についての困難度設定値は前掲表3-1に与えたものであり、受験者群500名についての能力設定値はそこで提示済みである。ＡＢＣそれぞれ

における受験者群 100 名、250 名、1,000 名に関する能力設定値の方法は 500 名についてのものと同様であり、その一覧を表 3-5 に掲示する。又、ＡＢＣデータ群作成への 01 データ生起は前述された方法で以って Excel 上で RAND 関数適用により出力されたものである。

**表 3-5　　ＡＢＣそれぞれにおける受験者群 100 名、250 名、1,000 名の能力設定値**

| | | | 100名 | 250名 | 1,000名 |
|---|---|---|---|---|---|
| A | 受験者能力設定値 | 最小値 | -2.876 | -3.244 | -3.200 |
| | | 標準正規分布 | N (0, 1) | N (0, 1) | N (0, 1) |
| | | 最大値 | 2.753 | 3.022 | 3.473 |
| B | 受験者能力設定値 | 最小値 | -2.876 | -3.244 | -3.200 |
| | | 標準正規分布 | N (0, 1) | N (0, 1) | N (0, 1) |
| | | 最大値 | 2.753 | 3.022 | 3.473 |
| C | 受験者能力設定値 | 最小値 | -3.000 | -3.000 | -3.000 |
| | | 刻み目 | 0.060 | 0.024 | 0.006 |
| | | 最大値 | 2.940 | 2.976 | 2.994 |

**表 3-6　　ＡＢＣそれぞれ 24 個の 01 データ上での指標値変動範囲**

| | 項目群 Cronbach's Alpha | 01データ 項目群ピアソン 主成分分析 第1主成分 寄与率 | 01データ 受験者群ピアソン 主成分分析 第1主成分 寄与率 | 01データ 項目群テトラコリック 主成分分析 第1主成分 寄与率 |
|---|---|---|---|---|
| A24データ指標値範囲 | 0.772 ｜ 0.969 | 10.6% ｜ 14.3% | 47.8% ｜ 51.0% | 18.7% ｜ 27.3% |
| B24データ指標値範囲 | 0.837 ｜ 0.978 | 14.6% ｜ 18.7% | 19.3% ｜ 27.9% | 19.4% ｜ 30.0% |
| C24データ指標値範囲 | 0.905 ｜ 0.988 | 24.1% ｜ 28.7% | 41.4% ｜ 43.8% | 35.5% ｜ 49.7% |

表 3-6 でのＡＢＣそれぞれにおける 24 データの項目群・受験者群第一主成分寄与率変動範囲は、受験者群 500 名に関する前掲表 3-2 から表 3-4 におけるＡＢＣそれぞれにおける項目群・受験者群第一主成分寄与率変動範囲とほぼ同等であり、狭小な変動域として示されている。それは、ＡＢＣそれぞれにおける 30 項目、60 項目、90 項目、120 項目、200 項目、250 項目の困難度分布設定値が同一形態とされ、ＡＢＣそれぞれにおける受験者群 100 名、250 名、500 名、1,000 名の能力分布設定値も同一形態とされていることによるものと推察される。又、受験者群 100 名、250 名、1,000 名に関するデータ群の Rasch 測定上での特徴は、受験者群 500 名について上述されたデータ群ＡＢＣの指標値比較まとめに準ずるものとなる筈である。表 3-6 においてもＣの 24 データについてのみに項目群ピアソン相関係数に基づく第一主成分寄与率として 20%を超える値が付されている。これは、ＡとＢに関する前掲表 3-2 と表 3-3 における 500 名データ群と比較して、Ｃに関する前掲表 3-4 での 500 名データ群における大きい能力推定値標準偏差と大きい項目群第一主成分寄与率との連関性が、Ｃでの受験者群 100 名、250 名、1,000 名に関するデータ群においても同様に観察されることの示唆である。

入学試験の様な実際の能力検定テストにおいては、素点得点に併せて Rasch 測定上での能力推定値は正規分布に準ずるものと想定される故に、その能力推定値標準偏差がＣに関する前掲表 3-4 に示された 1.80 前後に達することは一般的ではない様に推量される。従って、本節での模擬データ群による示唆の上では、現実 01 データにおいて項目間ピアソン相関係数に基づく項目群第一主成分寄与率が 20%を超えることは、難度上で高位あるいは低位への一方極端に密集する項目群からなる成るテスト

以外には、ほとんど起こり得ないことになる。但し、前章第5節で論証された様に、困難度推定値標準偏差が小さくなれば、受験者群第一主成分寄与率は小さくなることに留意される。更に、困難度推定値標準偏差が小さくなれば、困難度推定値についてのRasch測定妥当性・信頼性が低くなると上述により帰結される。

## 2．現実テスト

本節で分析するテストは、その構成概念を英語構文知識として、「代名詞」・「完了形」・「助動詞」・「態」・「不定詞」・「分詞」・「動名詞」・「比較」・「関係詞」・「仮定法」の文法範疇それぞれ10問前後から成る四肢選択100問(井澤[24]、2006, p. 3)を208問に拡大刷新したものである。2004年でのその作成時における狙いは、推測上で難易度範囲を広めた項目群の挿入と現実的な問題数増大により、Raschモデル適合度の高い項目群抽出可能性へのぼんやりとした淡い期待であった。そのテスト実施を、流通科学大学での筆者担当授業における2004年度後期開始時から2007年度後期開始時までの受験者総数506名(著者による研究考察使用承諾受領済み)に対して制限時間を正味95分としたものである。なお、受講者 30 人につきほぼ二人か三人の割合で存在した重複受験者群については第一回目の回答を採用して、未解答項目のある受験者群を除いた受験者総数が506名である。問題数が多過ぎるとの批判もあろうが、一項目一文を幹とする208行から成る英文への穴埋め四肢選択形式は制限時間 95 分内での応答時間の上で非常識ではない。応答速度テストの要素が含まれていたとしても、ほぼすべての受験者が制限時間内に項目全数への応答を終えていたことが一つの安心事である。

そのテスト結果の初期分析として、01データ素点上での3問への受験者群応答に関する点双列相関係数が負値となったことにより、その3問を当該受験者群への明らかな「不良項目」(静[25]、2002, pp. 276- 277、参照)として分析対象項目群から削除した。従って、本節での現実テスト分析提示は 506名による205項目への正誤反応データに基づいている。表3-7が、前節表提示と同じ要領で作成した現実205項目×506名データ指標値、前掲表3- 3における模擬一次元Bデータ群第5番(200項目×500名)、及び、10個の模擬一次元データ(205 項目×506 名)指標値平均を与えるものである。付加データ資料は、それぞれ上記の現実データ指標値と比較することを意図としている。特に、下段の資料付加は、現実テストにおける困難度推定値と能力推定値それぞれに関するRasch測定妥当性・信頼性の達成度を観る有用な指標となることを示すためである。それについては以下に詳述する。

現実データについて第一に注目されることは、項目群と受験者群のいずれにおいてもピアソン相関係数に基づく第一主成分寄与率が20%未満であり、項目群についてのテトラコリック相関係数に基づくその対応値も同様である。それは、前章第5節論証の例示であり、この現実データに関する困難度推定値と能力推定値いずれもの標準偏差が 1.00 前後の中庸な値であることに連関している。一方、Rasch測定上での困難度推定値と能力推定値いずれもの不変性成立度は決して低い値ではないと示されている。現実データの場合には、困難度推定値不変性成立度は受験者群能力一次元性充足度として、又、能力推定値不変性成立度は項目群内容一次元性充足度として理解される。表3-7は、現実取得正

誤反応データの項目群と受験者群の第一主成分寄与率がたとえ20%未満であるとしても、Rasch 測定上での上記それぞれの不変性成立程度確認の重要性を顕示している。その証左となるものが表 3-7 での中段と下段に与えた模擬一次元データ指標値である。

表 3-7　現実205項目×506名データ指標値、前掲表3- 3における模擬データ5(200項目×500名)、及び、10個の模擬一次元データ(205項目×506名)指標値平均

<table>
<tr><td rowspan="6">現実<br>データ<br>205項目<br>x<br>506名</td><td rowspan="6">01データ<br>項目群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>11.7%</td><td rowspan="6">01データ<br>受験者群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>19.0%</td><td rowspan="6">01データ<br>項目群テトラコリック<br>主成分分析<br>第1固有値寄与率<br>18.5%<br>GFI = 0.886</td><td colspan="13">Rasch 測定Facets出力項目困難度推定値</td></tr>
<tr><td>最小値</td><td>-4.45</td><td rowspan="2">標準偏差</td><td rowspan="2">1.22</td><td rowspan="2">範囲</td><td rowspan="2">7.65</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.99</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">204</td><td>ピアソン</td><td>0.853</td></tr>
<tr><td>最大値</td><td>3.20</td><td>ケンドールタウb</td><td>0.662</td></tr>
<tr><td colspan="13">Rasch 測定Facets出力受験者能力推定値</td></tr>
<tr><td>最小値</td><td>-1.05</td><td rowspan="2">標準偏差</td><td rowspan="2">0.86</td><td rowspan="2">範囲</td><td rowspan="2">5.58</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.96</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">506</td><td>ピアソン</td><td>0.839</td></tr>
<tr><td>最大値</td><td>4.53</td><td>ケンドールタウb</td><td>0.655</td></tr>
<tr><td rowspan="6">Bデータ群<br>No. 5<br>(200項目<br>x<br>500名)</td><td rowspan="6">01データ<br>項目群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>14.6%</td><td rowspan="6">01データ<br>受験者群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>20.5%</td><td rowspan="6">01データ<br>項目群テトラコリック<br>主成分分析<br>第1固有値寄与率<br>23.2%<br>GFI = 0.945</td><td colspan="13">Rasch 測定Facets出力項目困難度推定値</td></tr>
<tr><td>最小値</td><td>-3.43</td><td rowspan="2">標準偏差</td><td rowspan="2">1.22</td><td rowspan="2">範囲</td><td rowspan="2">7.18</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.99</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">200</td><td>ピアソン</td><td>0.925</td></tr>
<tr><td>最大値</td><td>3.75</td><td>ケンドールタウb</td><td>0.777</td></tr>
<tr><td colspan="13">Rasch 測定Facets出力受験者能力推定値</td></tr>
<tr><td>最小値</td><td>-3.26</td><td rowspan="2">標準偏差</td><td rowspan="2">0.99</td><td rowspan="2">範囲</td><td rowspan="2">6.04</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.97</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">500</td><td>ピアソン</td><td>0.888</td></tr>
<tr><td>最大値</td><td>2.78</td><td>ケンドールタウb</td><td>0.707</td></tr>
<tr><td rowspan="6">10個の<br>模擬<br>一次元<br>205x506<br>データ<br>平均値</td><td rowspan="6">01データ<br>項目群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>10.9%</td><td rowspan="6">01データ<br>受験者群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>18.7%</td><td rowspan="6">01データ<br>項目群テトラコリック<br>主成分分析<br>第1固有値寄与率<br>17.1%<br>GFI = 0.890</td><td colspan="13">Rasch 測定Facets出力項目困難度推定値</td></tr>
<tr><td>最小値</td><td>-4.89</td><td rowspan="2">標準偏差</td><td rowspan="2">1.24</td><td rowspan="2">範囲</td><td rowspan="2">8.15</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.99</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">204</td><td>ピアソン</td><td>0.925</td></tr>
<tr><td>最大値</td><td>3.27</td><td>ケンドールタウb</td><td>0.761</td></tr>
<tr><td colspan="13">Rasch 測定Facets出力受験者能力推定値</td></tr>
<tr><td>最小値</td><td>-1.26</td><td rowspan="2">標準偏差</td><td rowspan="2">0.89</td><td rowspan="2">範囲</td><td rowspan="2">6.11</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.96</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">506</td><td>ピアソン</td><td>0.858</td></tr>
<tr><td>最大値</td><td>4.85</td><td>ケンドールタウb</td><td>0.662</td></tr>
</table>

ＡあるいはＣからではなく、Ｂにおいて当該現実データの項目数と受験者数にほぼ相応する第 5 番データを表 3-7 に含めたのは、その模擬データの項目群困難度と受験者群能力いずれもの正規分布設定値によりもたらされたその Rasch 測定上での両群母数推定値分布正規性への目視確認の上でのことである。当然ながら、当該現実データの両群母数推定値分布正規性が観察される故に、両データに関する指標値比較が有益となることを期待した上での上記模擬一次元データ作成であった訳である。模擬データが現実データに指標値の上で全般的に僅かばかり上回っているが、見事にその期待がかなえられて、両者の同等性が顕著に観察される。模擬データが有している完全な一次元性からの判断により、その項目群困難度と受験者群能力についての母数推定値不変性成立度と大差のないそれぞれの対応値が現実データにおいて示されていることは、この現実データの低いとはいえない一次元充足度の証左一端である。更に、模擬データにおける困難度と能力の推定値標準偏差が現実データでの各対応値とほぼ同等であり、それぞれ 1.00 前後の値として与えられている。従って、現実データと模擬データはそれぞれ一例ではあるが、現実データにおいて Rasch 測定上での両母数の推定値分布正規性が観測されて、それぞれの標準偏差が 1.00 前後であるならば、現実データにおいて項目群と受験者群いずれもの第一主成分寄与率が 20%を大きく超えることは一般的に起こり得ないと推量される。

上記推量に関して上の模擬データ一例による議論を更に充分といえる程度に補完するものが表 3-7 の下段に与えた資料指標値である。その作成は、前節での模擬一次元データと同じく、現実テストの Rasch 測定により得られた各項目困難度推定値と各受験者能力推定値の一対すべての組み合わせにおける Excel 上での Rasch 項目分析モデル期待正答確率(前章第一節掲示式(2))出力値への RAND 関数適用による 01 データ生起に基づいている。その 10 個のデータを生成した上でその指標値平均がとられている。Rasch 測定母数推定値不変性成立度として示される Rasch 測定妥当性・信頼性に関するそ

の指標値平均は、現実データの項目数と困難度推定値分布、及び、受験者数と能力推定値分布によって「固有とされる関係枠」(Rasch[26], 1977, p. 77、参照)の下で、現実データから Rasch モデルにより構成された最大限に一次元的な尺度構成体系(Linacre[13], 1998b, p. 267、参照)の上での近似最大値と理解される。従って、当該現実データの困難度推定値と能力推定値についての不変性成立度をそれぞれ 10 個の模擬一次元データ平均値として示された対応値で除することにより、当該現実データに関する一次元尺度構成体系上でのその達成度指標が与えられる。その指標値は、205 項目困難度推定値に関しては、0.853 / 0.925 ≒ 0.922、0.662 / 0.761 ≒ 0.870、並びに、506 名能力推定値に関しては、0.839 / 0.858 ≒ 0.978、0.655 / 0.662 ≒ 0.989 と算出される。各一対での小値を採用して、当該現実データの 205 項目と 506 名のそれぞれに関する一次元 Rasch 測定尺度構成体系の達成度指標値は 0.87 と 0.98 である。なお、前者の 87%が当該現実データにおける受験者群に関する一次元尺度構成体系達成度であり、後者の 98%が項目群に関する一次元尺度構成体系達成度である。これにより、当該現実データの一次元尺度構成体系達成度は大変に高いとみなされる。但し、その指標は、前節での模擬一次元データ群で明確に観察された様に、項目数と受験者数の大きさにも依存していることに留意される。又、10 個の模擬一次元データ平均値としての項目群・受験者群第一主成分寄与率はいずれも 20%未満の値であり、その一次元性充足度指標値としての下限臨界値 20%の妥当性欠如が再確認される。なお、表 3-7 での項目群困難度と受験者群能力それぞれの推定値不変性成立度の指標値算出の上で、受験者群と項目群それぞれについての Rasch 測定標準化残差主成分分析に基づく負荷量記号正負分別法のデータ折半法における下限信頼精度優位性(平越・井澤[11]、2008, p. 77、参照)を現実 205 項目×506 名データのみに関して表 3-8 に示しておく。

表 3-8　データ折半法による両母数推定値不変性成立度の異同一例

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 正の負荷量群 対 負の負荷量群 | 項目困難度推定値不変性 (受験者群一次元性) | 正の負荷量268名 対負の負荷量238名 | 有効数 | 204 | ピアソン | 0.853 |
| | | | | | | ケンドールタウb | 0.662 |
| | | 受験者能力推定値不変性 (項目群一次元性) | 正の負荷量111項目 対負の負荷量94項目 | 有効数 | 506 | ピアソン | 0.839 |
| | | | | | | ケンドールタウb | 0.655 |
| 2 | 推定値上位群 対 推定値下位群 | 項目困難度推定値不変性 (受験者群一次元性) | 受験者能力推定値 上位259名対下位247名 | 有効数 | 204 | ピアソン | 0.896 |
| | | | | | | ケンドールタウb | 0.720 |
| | | 受験者能力推定値不変性 (項目群一次元性) | 項目困難度推定値 上位102項目対下位103項目 | 有効数 | 505 | ピアソン | 0.850 |
| | | | | | | ケンドールタウb | 0.701 |
| 3 | 前半部分群 対 後半部分群 | 項目困難度推定値不変性 (受験者群一次元性) | 前半部253名 対後半部253名 | 有効数 | 205 | ピアソン | 0.979 |
| | | | | | | ケンドールタウb | 0.896 |
| | | 受験者能力推定値不変性 (項目群一次元性) | 前半部102項目 対後半部103項目 | 有効数 | 506 | ピアソン | 0.891 |
| | | | | | | ケンドールタウb | 0.717 |
| 4 | 奇数番号群 対 偶数番号群 | 項目困難度推定値不変性 (受験者群一次元性) | 奇数番号253名 対偶数番号253名 | 有効数 | 205 | ピアソン | 0.976 |
| | | | | | | ケンドールタウb | 0.892 |
| | | 受験者能力推定値不変性 (項目群一次元性) | 奇数番号103項目 対偶数番号102項目 | 有効数 | 505 | ピアソン | 0.936 |
| | | | | | | ケンドールタウb | 0.780 |
| 5 | 無作為二群分別 | 項目困難度推定値不変性 (受験者群一次元性) | 無作為抽出253名 対その残り253名 | 有効数 | 205 | ピアソン | 0.980 |
| | | | | | | ケンドールタウb | 0.886 |
| | | 受験者能力推定値不変性 (項目群一次元性) | 無作為抽出102項目 対その残り103項目 | 有効数 | 505 | ピアソン | 0.909 |
| | | | | | | ケンドールタウb | 0.757 |

Rasch 測定の上での一つの属性として、同一受験者群に対する全体項目群データと部分項目群データの同一項目群に関する困難度推定値間の相関係数は、両群データにおける同一項目群に関する各素点得点の同一性とその測定上での十分統計量としての性質により、常に 1.00 となる特性が存在する (Smith, Jr.[16], 2002, p. 206)。その特性は、筆者によっても確認済みである。従って、その特性との関

係で以って、全体項目群データから Rasch モデル適合度の高い項目群を抽出することが可能となる。その方法を以下に述べる。

平越・井澤[11](2008)により、Rasch 測定上での項目群困難度と受験者群能力のいずれもの母数推定値が広範囲に分布していなければ、その Rasch 測定妥当性に欠けると示されている。それは、項目群困難度と受験者群能力のいずれもの推定値分布における大きな標準偏差と高い Rasch 測定妥当性の連関性を意味している。データの Rasch 測定後に、受験者群を選別してその能力推定値分布の標準偏差を高めることは測定目的の上で当然に不能である。然しながら、テスト構成概念研究考察の上では、テスト受験者群母集団に対して普遍的に有用な項目群に関する特性の実体認知が求められて、Rasch モデル適合度の高い項目群の収集・集積の必要性もある。困難度推定値が広範囲に分布して、更に、前掲表 3-2 から表 3-4 において視認される様に、その困難度推定値が正規分布よりも一様分布に近づくならば、その標準偏差は大きくなる。従って、データの Rasch 測定後に、一方法として、困難度推定値の最大値と最小値の範囲内で最大限に等間隔性を保つ項目群を抽出すれば、その項目群の困難度推定値標準偏差は元データ全体項目群のそれよりも一般的に高くなる。困難度推定値等間隔性が高い程度に保持された抽出項目群から成るデータの Rasch 測定困難度推定値は、上記のモデル特性により、元データにおけるその抽出項目群の等間隔的推定値との同一順序性と等間隔性が最大限に保持された一様分布となる。能力検定テストにおける受験者群能力は一般的に正規分布に準ずる故に、前掲表 3-2 におけるＡデータ群について観察される様に、モデル適合度の高い項目群としての高い困難度推定値不変性成立度が達成される。以下に、その一例を与えて、そのデータに関する指標値を観察する。

**表 3-9　現実抽出 30 項目×505 名データ指標値、前掲表 3-2 における模擬データ 1(30 項目×500 名)、及び、10 個の模擬一次元データ(30 項目×505 名)指標値平均**

<table>
<tr><td rowspan="6">現実<br>データ<br>抽出<br>30項目<br>x<br>505名</td><td rowspan="6">01データ<br>項目群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>12.9%</td><td rowspan="6">01データ<br>受験者群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>43.8%</td><td rowspan="6">01データ<br>項目群テトラコリック<br>主成分分析<br>第1固有値寄与率<br>22.7%<br>GFI = 0.810</td><td colspan="13">Rasch測定Facets出力項目困難度推定値</td></tr>
<tr><td>最小値</td><td>-4.44</td><td rowspan="2">標準偏差</td><td rowspan="2">2.25</td><td rowspan="2">範囲</td><td rowspan="2">8.02</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.99</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">30</td><td>ピアソン</td><td>0.913</td></tr>
<tr><td>最大値</td><td>3.58</td><td>ケンドールタウb</td><td>0.849</td></tr>
<tr><td colspan="13">Rasch測定Facets出力受験者能力推定値</td></tr>
<tr><td>最小値</td><td>-1.83</td><td rowspan="2">標準偏差</td><td rowspan="2">1.10</td><td rowspan="2">範囲</td><td rowspan="2">6.80</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.77</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">476</td><td>ピアソン</td><td>0.389</td></tr>
<tr><td>最大値</td><td>4.97</td><td>ケンドールタウb</td><td>0.292</td></tr>
<tr><td rowspan="6">Aデータ群<br>No. 1<br>(30項目<br>x<br>500名)</td><td rowspan="6">01データ<br>項目群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>13.4%</td><td rowspan="6">01データ<br>受験者群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>51.0%</td><td rowspan="6">01データ<br>項目群テトラコリック<br>主成分分析<br>第1固有値寄与率<br>25.6%<br>GFI = 0.896</td><td colspan="13">Rasch測定Facets出力項目困難度推定値</td></tr>
<tr><td>最小値</td><td>-4.22</td><td rowspan="2">標準偏差</td><td rowspan="2">2.57</td><td rowspan="2">範囲</td><td rowspan="2">8.35</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">1.00</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">30</td><td>ピアソン</td><td>0.929</td></tr>
<tr><td>最大値</td><td>4.13</td><td>ケンドールタウb</td><td>0.791</td></tr>
<tr><td colspan="13">Rasch測定Facets出力受験者能力推定値</td></tr>
<tr><td>最小値</td><td>-4.02</td><td rowspan="2">標準偏差</td><td rowspan="2">1.22</td><td rowspan="2">範囲</td><td rowspan="2">7.62</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.79</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">500</td><td>ピアソン</td><td>0.581</td></tr>
<tr><td>最大値</td><td>3.60</td><td>ケンドールタウb</td><td>0.468</td></tr>
<tr><td rowspan="6">10個の<br>模擬<br>一次元<br>30x505<br>データ<br>平均値</td><td rowspan="6">01データ<br>項目群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>14.7%</td><td rowspan="6">01データ<br>受験者群ピアソン<br>主成分分析<br>第1固有値<br>寄与率<br>43.9%</td><td rowspan="6">01データ<br>項目群テトラコリック<br>主成分分析<br>第1固有値寄与率<br>26.7%<br>GFI = 0.888</td><td colspan="13">Rasch測定Facets出力項目困難度推定値</td></tr>
<tr><td>最小値</td><td>-4.85</td><td rowspan="2">標準偏差</td><td rowspan="2">2.37</td><td rowspan="2">範囲</td><td rowspan="2">8.64</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.99</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">30</td><td>ピアソン</td><td>0.916</td></tr>
<tr><td>最大値</td><td>3.79</td><td>ケンドールタウb</td><td>0.796</td></tr>
<tr><td colspan="13">Rasch測定Facets出力受験者能力推定値</td></tr>
<tr><td>最小値</td><td>-2.64</td><td rowspan="2">標準偏差</td><td rowspan="2">1.26</td><td rowspan="2">範囲</td><td rowspan="2">7.76</td><td rowspan="2">推定値分離信頼性係数</td><td rowspan="2">0.81</td><td rowspan="2">推定値不変性</td><td rowspan="2">有効数</td><td rowspan="2">493</td><td>ピアソン</td><td>0.566</td></tr>
<tr><td>最大値</td><td>5.12</td><td>ケンドールタウb</td><td>0.428</td></tr>
</table>

表 3-9 が、表 3-7 での現実 205 項目×506 名データの Rasch 測定上で 205 項目に付された困難度推定値の上で 0.225 間隔基準により抽出した 30 項目×505 名データについての指標値を与えるものである。併せて、前掲表 3-2 における模擬一次元Ａデータ群第 1 番(30 項目×500 名)、及び、表 3-7 での 10 個の模擬一次元データ(205 項目×506 名)と同じ要領で作成した 10 個の模擬一次元データ(30 項目

×505 名)の指標値平均を与えている。なお、上段の現実 30 項目×505 名データについては、元データ受験者数は 506 名であるが、30 項目抽出による満点受験者 1 名を削除している。

前述の模擬一次元 A 群第 1 番データでの 30 項目困難度推定値標準偏差 2.57 とほぼ同程度の大きな値 2.25 が現実抽出 30 項目×505 名データの項目困難度推定値標準偏差として与えられている。その大きな標準偏差に連関して、受験者群第一主成分寄与率としての大きな値 43.8%が付されて、困難度推定値不変性成立度(受験者群能力一次元性充足度)も大変に高い値となっている。その値は、10 個の模擬一次元データ平均値におけるその対応値に匹敵する高さであり、505 名の受験者群母集団におけるこの現実 30 項目の困難度推定値不変性・信頼性・再現性・同一順序性の高さが窺われる。なお、少ない項目数の影響により、受験者群 505 名に関する能力推定値不変性成立度が低い値として示され、10 個の模擬一次元データ平均値におけるその対応値との比較の上で、現実 30 項目についての上述された一次元尺度構成体系達成度は 0.389 / 0.566 ≒ 0.687、0.292 / 0.428 ≒ 0.682 として与えられる。又、その項目群折半上での 476 名能力推定値同等性検証については、z 値の絶対値として 2 を超える値が付された受験者数の割合は 34 / 476 ≒ 0.071 と算出される。従って、Facets 出力上での能力推定値標準誤差への参照を含めれば、90%有意水準の上で 476 名能力推定値同等性が認められることになる(Bond and Fox[15], 2001, pp. 56–57、参照)。但し、重要事項として、その 30 項目抽出元の 205 項目についての同質内容項目群無限母集団を想定すれば、困難度推定値等間隔性に基づく項目群抽出はこの 30 項目のみに限られないことに留意される。更に、現実抽出 30 項目×505 名データにおける受験者群能力推定値の高い分布正規性が視認され、その標準偏差として示された大きくはない値 1.10 に相応して、その項目群第一主成分寄与率としての低い値 12.9%が観察される。

## Ⅳ. おわりに

本稿での論証と例証により、正誤反応データ一次元性充足度の判断基準値としての項目群第一主成分(固有値)寄与率下限臨界値 20%は妥当性に欠けることが明白である。この点のみにおいても、取得 01 データの Rasch 項目分析モデル適用の上での項目群と受験者群に関する母数推定値不変性成立度確認有用性が認められる。又、著者の知り得る限り、人間科学研究考察上での主成分・因子分析は一般的に項目群のみを対象としているが、項目群だけでは測定尺度構成についての考察は不十分である。項目群の主成分・因子分析は項目群一次元性充足度(受験者群能力推定値不変性成立度)を観るものではあるが、潜在特性(受験者能力)測定尺度構成上での有用な項目群は、項目群一次元性に関する指標ではなく、受験者群能力一次元性充足度(項目群困難度推定値不変性成立度)に関する指標によって明確とされる。従って、01 データ上での主成分・因子分析に限るならば、受験者群についての主成分・因子分析が項目群に関する素点得点の信頼性・再現性・有用性への判断情報を与えることになる。それが、前節表 3-9 において簡明直截に示されている。なお、第一主成分寄与率としての下限臨界値 20%に妥当性はないとしても、項目群・受験者群第一主成分寄与率の項目数と受験者数に影響されない高

い安定性と Rasch 測定妥当性・信頼性指標値の高さとの連関性が前章において観察される。それは、正誤反応データの第一主成分寄与率への参照は無益ではないとの例示一端である。

表 4-1　本稿中心論点連関性

| 項目群 | 困難度推定値標準偏差 小 | 困難度推定値標準偏差 大 | 受験者群 | 能力推定値標準偏差 小 | 能力推定値標準偏差 大 |
|---|---|---|---|---|---|
| 項目群第一主成分寄与率 | 大 | 小 | 受験者群第一主成分寄与率 | 大 | 小 |
| 受験者群第一主成分寄与率 | 小 | 大 | 項目群第一主成分寄与率 | 小 | 大 |
| 困難度推定値不変性成立度 | 小 | 大 | 能力推定値不変性成立度 | 小 | 大 |
| 困難度推定値妥当性・信頼性 | 小 | 大 | 能力推定値妥当性・信頼性 | 小 | 大 |

最後に、本稿での中心的な論点の連関性をまとめとして表 4-1 に与え、第一主成分寄与率は、項目群と受験者群においてトレードオフの関係にあることを明示する。各第一主成分寄与率の大きさを目安に、項目群と受験者群それぞれについての一次元性を高めることには意義もあるが、その一方的固執は偏向であろう。特に項目群第一主成分寄与率の上昇だけに注目すれば、設定項目群の困難度を同程度に揃え、狭小な困難度分布幅に設計すれば事は済む。しかし、それはテスト本来の目的・意味の履き違えである。つまり、明確なテスト目的で以って、項目群と受験者群の相互関係への配慮からバランスの良いテスト設計を行うことが肝要であり、第一主成分寄与率の単独指標のみならず、本稿での提示を含む多くの参照可能な指標を用いた総合的判断に基づくテスト設計・解析が重要なのである。

**参照文献**


1) M. D. Reckase. 1979. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 3, pp. 207- 230.

2) 静　哲人　2007. 『基礎から深く理解するラッシュモデリング』　大阪：　関西大学出版部

3) G. Rasch. 1960. *Probabilistic models for some intelligence and attainment tests.* The Danish Institute for Educational Research. (Reprinted in 1980 by the University of Chicago Press with a Foreword and Afterword by Wright, B. D.)

4) G. Rasch. 1961. On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Theory of Probability, Vol. IV* (pp. 321- 333). Berkeley: University of California Press.

5) B. D. Wright. 1992. IRT in the 1990s: Which models work best? *Rasch Measurement Transactions*, 6, 1 in Linacre, J. M. (Ed.), 1996, *Rasch Measurement Transactions, Part 2* (pp. 196- 200). Chicago: MESA Press.

6) 芝　祐順編著　1991. 『項目反応理論　―　基礎と応用』　東京大学出版会

7) 渡辺直・野口裕之編著　1999. 『組織心理測定論　―　項目反応理論のフロンティア』　東京：白桃書房

8) 豊田秀樹編著　2002. 『項目反応理論[事例編]』　東京：　朝倉書店

9) J. M. Linacre. 1998a. Rasch first or factor first? *Rasch Measurement Transactions*, 11, 4, p. 603.

10) B. D. Wright. 1996. Local dependency, correlations and principal components. *Rasch Measurement Transactions*, 10, 3, 509- 511.

11) 平越裕之・井澤廣行 2008.「Rasch 項目分析モデル測定母数推定値分布幅、テスト次元性、並びに、Rasch 測定適用妥当性」『流通科学大学論集 — 流通・経営編』第 20 巻、第 2 号、pp. 61-85.

12) J. M. Linacre. 1999a. Category disordering vs. step disordering. *Rasch Measurement Transactions*, 13, 1, 675- 678.

13) J. M. Linacre. 1998b. Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2, 3, 266- 283.

14) J. M. Linacre. 1999b. Explorations into local independence with T-Rasch. *Rasch Measurement Transactions*, 13, 3, 710.

15) T. G. Bond & C. M. Fox. 2001. *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

16) E. V. Smith, Jr. 2002. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 2, 205- 231.

17) A. Tennant & J. F. Pallant. 2006. Unidimensionality matters. *Rasch Measurement Transactions*, 20, 1, 1048- 1051.

18) B. D. Wright & M. H. Stone. 2004. *Making measures.* Chicago: The Phaneron Press, Inc.

19) D. Andrich. 1982. An Index of Person Separation in Latent Trait Theory, the Traditional KR. 20 Index, and the Guttman Scale Response Pattern. *Educational Research and Perspectives*, 9, 95- 104.

20) C. M. Fox & J. A. Jones. 1998. Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45, 1, 30- 45.

21) W. Zhu, W. F. Updyke, & C. Lewandowski. 1997. Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement*, 1, 4, 286- 304.

22) J. M. Linacre. 1989- 2001. *A user's guide to FACETS: Rasch measurement computer program.* Chicago: Winsteps.com.

23) B. D. Wright & G. N. Masters. 1982. *Rating scale analysis.* Chicago: MESA Press.

24) 井澤廣行 2006. 「Rasch モデル母数推定値同等性検証のためのカイ二乗検定有用性」『流通科学大学論集 — 人間・社会・自然編』第 18 巻、第 3 号、pp. 1- 14.

25) 静 哲人 2002. 『英語テスト作成の達人マニュアル』 東京: 大修館書店

26) G. Rasch. 1977. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In Blegvad, M. (Ed.), *The Danish Yearbook of Philosophy*, (pp. 58- 94). Copenhagen: Munksgaard.