

書評

静 哲人著『基礎から深く理解するラッシュモデリング』

関西大学出版部 2007年3月31日刊行 373頁

ISBN 978-4-87354-447-2 C3033

井澤 廣行*

Hiroyuki Izawa

Rasch モデル、特に、01 データに適用する Rasch 項目分析モデルの測定理論についての基本理解に資する良書である。読者対象はテスト分析専攻学徒に限られない。現代において順序尺度データへの有用な測定理論の一つである Rasch モデルについての基本論説が主題ではあるが、テスト分析においてその対処上での知識と運用技能が求められる統計基礎数理と古典的テスト理論についても説明されている。従って、構成は、「第1部: 数学的準備」、「第2部: テスト理論の基礎」、及び、「第3部: ラッシュ測定理論」とされている。言葉は平易であり、現実テスト分析に際しての必須適用手段である Excel が駆使されて、その出力図表で以って読者理解の促進と深化が意図されている。

著者の聡明さが随所に窺われる。評者が明確に認識していなかった次なるものはその一例である。『[Rasch 測定上での]ある項目の弁別力とは、実は単一の数値ではなく、 θ_n [受験者 n の能力母数]の関数である』(p. 280)。従って、Rasch 項目分析モデル測定上での項目弁別力は、01 データにおける項目弁別力として単一数値により示される弁別指標(静・竹内・吉澤¹⁾、2002, pp. 71-72、参照)ないしは点双列相関係数(静・竹内・吉澤¹⁾、2002, p. 132、参照)とは実体概念に違いがあると理解される。Rasch 測定プログラム上でモデル適合度指標として一般的に出力される Infit・Outfit の χ^2 平均平方統計量は各項目と各受験者それぞれに関する Rasch モデル上での弁別力に大きく依拠している。その Infit・Outfit 統計量と 01 データ点双列相関係数との相関は常に高いとは限らないと平越・井澤²⁾(2008)により示されており、その実体としての違いが窺われる。

上記の Infit・Outfit モデル適合度統計量上での標本の大きさに即応する不適合基準値について、著者により、『機械的あるいは絶対的なルールは存在しない』(p. 319)と指摘されている。これは、著者によっては参照されていない Karabatsos³⁾(2000)の精緻で的確な例証に基づく Infit・Outfit 統計量モデル不適合度検知限界性への言及からもうなずける。然しながら、著者

* 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

(2007年9月4日受理)

によるその言葉の前に提示されている表 12.11 (p. 318) に評者は僅かにしても懸念を抱く。その表は、「サンプルサイズによる平方平均の許容限界値の変化」と題され、Smith, Schumacker, and Bush⁴⁾ (1998) への参照に基づいて著者により作成された Infit と Outfit の χ^2 平均平方統計量それぞれについての標本の大きさに準ずるモデル不適合基準値である。著者によるその表作成法は評者に不詳ではあるが、Smith, Schumacker, and Bush⁴⁾ (1998) による模擬データ作成は、標本の大きさに係わらず、受験者能力推定値の標準正規分布設定に併せて、項目困難度推定値は狭い幅で -2.0 から 2.0 に渡る一様分布に基づくものとされている (p. 70)。同時最尤推定法による母数推定値の狭小な分布幅は、本質的に、データ多次元性ないしは Rasch 測定適用妥当性欠如を示唆しており (平越・井澤²⁾, 2008)、Smith, Schumacker, and Bush⁴⁾ (1998) の作成模擬データ群自体に Rasch モデル適合度において少なからずの変動性が存在していると推測される。従って、標本の大きさを同一とするデータ群において Infit・Outfit χ^2 平均平方統計量の分布平均値がとらわれているとしても、標本の大きさが異なるデータ群の間にはその平均値の上でもモデル適合度が一定であるとは想定され難い。Smith, Schumacker, and Bush⁴⁾ (1998) のそのデータ統計量に基づく著者作成の表 12.11 は、結果として、モデル適合度を一定値とする基準のもとでの標本の大きさに対応する Infit・Outfit χ^2 平均平方統計量の「許容限界値の変化」を示すものとはならない。その表を「許容限界値の変化」傾向の一例と理解することが正しく、Rasch 測定適用者がその表を鵜呑みにして取得データにおける各項目と各受験者についてのモデル適合度を一律に判断するべきではない。『Rasch モデル適合度統計量は[得られた]データと同様に局部的そして一時的である』と Wright and Stone⁵⁾ (2004, p. 33) によって指摘されている。標本の大きさのみならず個別データの特異性に依存する Infit・Outfit 統計量の変動性については留意が必要である。それが、その Rasch モデル適合度統計量上での「許容限界値」についての著者による『機械的あるいは絶対的なルールは存在しない』(p. 319) との言葉の意味である。

著者により、Linacre⁶⁾ (1998a) への参照の上で、『残差を主成分分析にかけて、意味のある第 1 主成分(あるいは因子)が取り出せないことの確認を行うことが望ましいとされます』(p. 366) と述べられている。この方法は、Wright⁷⁾ (1996) によって、順序尺度素点データへの因子分析はその適用妥当性に欠ける (p. 509) とされた上で、データ局所独立性充足程度の確認法として Rasch 測定ロジット残差主成分分析の有用性が指摘されたことに由来する。その後、Wright⁷⁾ (1996) が唱えたその Rasch 測定ロジット残差主成分分析に併せて残差そのものの主成分分析よりも分析結果明瞭性の点で僅かにしても優る Rasch 測定標準化残差主成分分析の適用有用性がデータ次元性充足度検知法として Linacre⁶⁾ (1998a) により実証された。その理論的適用根拠が更に言及されて (Linacre⁸⁾, 1998b, p. 636)、その方法は、現在では識者によりデータ局所独立性充足度確認法というよりもむしろ特異項目機能 (DIF) 察知を含めた包括的データ次元性検証法として位置付けられている (Bond and Fox⁹⁾, 2001, pp. 155-156; Smith, Jr.¹⁰⁾, 2002; Tennant and Pallant¹¹⁾,

2006; Wright and Stone⁵⁾, 2004, p. 21)。その Rasch 測定標準化残差主成分分析有用性についての評者の理解は次の通りである (Linacre⁸⁾, 1998b, p. 636、参照)。

Rasch 測定標準化残差は Rasch モデルでは説明されない 01 素データに内在する構成分であり、その項目群主成分分析による第 1 主成分は Rasch 測定誤差の主たる合成複合物であると考えられる。その第 1 主成分負荷量はデータ Rasch 測定誤差の項目群に関する主たる複合物とのピアソン相関係数となり、その負荷量正負記号分別は Rasch モデルでは説明されない項目群構成分の二極分化を示す。「データ次元性」により規定される (Wright¹²⁾, 1991, p. 158) Rasch モデルでは説明されない項目群構成分とは、項目群全体としての多次元性総量と解釈されるべきものと判断されて、その二極分化は項目群異質性の分化に他ならない。従って、Rasch 測定標準化残差の項目群主成分分析により出力される第 1 主成分負荷量の正負記号分別は Rasch 測定上での項目群異質性二極分化への指標となる。Rasch 測定標準化残差の受験者群についての主成分分析についても同様である。

上記により、Rasch 測定標準化残差の項目群と受験者群それぞれについて主成分分析により得られる第 1 主成分負荷量の正負記号分別がそれぞれに関する異質性二極分化への最大限的明示指標と捉えられる。従って、上に引用した著者による言葉 (p. 366) の後に追加参照されるべきことは、Rasch 測定理論の真髄である「不変性」(Andrich, 1988¹³⁾, pp. 21-22, and 2005¹⁴⁾, p. 53) のデータ充足度についての有用性の高い一つの検証法である。それは、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おける負荷量記号の正負分別に基づく受験者群折半と項目群折半 (Linacre⁸⁾, 1998b; Wright⁷⁾, 1996; Wright and Stone⁵⁾, 2004, pp. 19-21、参照) による下位データ間での項目群と受験者群それぞれにおける母数推定値不変性成立程度の確認である。データ次元性を意味するテスト構成概念妥当性の判断のためには、その方法が各項目と各受験者について個別に付される Infit・Outfit モデル適合度統計量に優るとの評者の見解である (平越・井澤²⁾, 2008、参照)。更に、項目群と受験者群それぞれに関するその折半法に基づく下位データ間に算出される二分標本 t 値 (Smith¹⁵⁾, 1996; Smith and Suh¹⁶⁾, 2003, p. 155; Wright and Stone¹⁷⁾, 1979, pp. 94-95) ないしは χ^2 検定 (McNamara¹⁸⁾, 1991, pp. 153-154) による母数推定値同等性の統計的検証が、母数推定値不変性確認補完のために望まれる。

以上少々詳しくすぎる程に、著者の言葉 (p. 319、及び、p. 366) に注釈を加えた。これは、Rasch モデル測定適用者が考慮すべきことであると評者が正にその適用経験の上で認識しているからである。これは、勿論、本書での読者に資する Rasch モデル理解への著者による指針提示に対する評者敬意の程を下げるものでは決してない。本書が日本でのテスト分析専攻学徒への必須基本図書の一つになることに疑いはない。

【参考文献】

- 1) 静 哲人・竹内 理・吉澤清美 2002. 『外国語教育リサーチとテストの基礎概念』 大阪： 関西大学出版部
- 2) 平越裕之・井澤廣行 2008(予定). 「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、頁未詳
- 3) G. Karabatsos. 2000. A critique of Rasch residual fit statistics. *Journal of Applied Measurement*, 1, 2, pp. 152–176.
- 4) R. M. Smith, R. E. Schumacker, & M. J. Bush. 1998. Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement*, 2, 1, pp. 66–78.
- 5) B. D. Wright & M. H. Stone. 2004. *Making measures*. Chicago: The Phaneron Press, Inc.
- 6) J. M. Linacre. 1998a. Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2, 3, pp. 266–283.
- 7) B. D. Wright. 1996. Local dependency, correlations and principal components. *Rasch Measurement Transactions*, 10, 3, pp. 509–511.
- 8) J. M. Linacre. 1998b. Structure in Rasch residuals: Why principal components analysis? *Rasch Measurement Transactions*, 12, 2, p. 636.
- 9) T. G. Bond & C. M. Fox. 2001. *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- 10) 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, pp. 205–231.
- 11) A. Tennant & J. F. Pallant. 2006. Unidimensionality matters. *Rasch Measurement Transactions*, 20, 1, pp. 1048–1051.
- 12) B. D. Wright. 1991. Scores, reliabilities and assumptions. *Rasch Measurement Transactions*, 5, 3 in Linacre, J. M. (Ed.), 1995, *Rasch Measurement Transactions, Part 1* (pp. 157–158). Chicago: MESA Press.
- 13) D. Andrich. 1988. *Rasch models for measurement*. Newbury Park, California: Sage Publications, Inc.
- 14) D. Andrich. 2005. The Rasch model explained. In Alagumalai, S, Curtis, D. D., & Hungi, N (Eds.), *Applied Rasch measurement: A book of exemplars* (pp. 27–59). The Netherlands: Springer.
- 15) R. M. Smith. 1996. A comparison of the Rasch separate calibration and between-fit methods of detecting item bias. *Educational and Psychological Measurement*, 56, 1, pp. 403–418.
- 16) R. M. Smith & K. K. Suh. 2003. Rasch fit statistics as a test of the invariance of item parameter estimates. *Journal of Applied Measurement*, 4, 2, pp. 153–163.
- 17) B. D. Wright & M. H. Stone. 1979. *Best test design*. Chicago: MESA Press.
- 18) T. McNamara. 1991. Test dimensionality: IRT analysis of an ESP listening test. *Language Testing*, 8, 2, pp. 139–159.