

初年次基礎学力確認テスト成績と 中間・期末試験成績の比較

齊 藤 善 弘

概 要

本稿では現代テスト理論の一つである項目反応理論を数学基礎学力確認テスト結果への適用を試みる。そして、古典的テスト理論による結果との比較、および数理授業科目の中間・期末試験成績との比較を行う。

1. はじめに

岐阜聖徳学園大学経済情報学部では平成28年度から入学直後に基礎学力確認テストを実施している。これは入学者の学生の能力を把握し、学生に合わせた授業内容の提供が主な目的である。現在、多様な学生の受け入れを行っている大学が多く存在し、筆記試験を課す大学もあれば、推薦入試の面接のみで入学する学生もいる。特に「数学」に関して言えば、高校で必修科目の「数学Ⅰ」のみを修得してくる学生もいれば、理系で必須とされる「数学Ⅲ」まですべての数学科目を修得してくる学生もいる。また「数学Ⅰ」のみ修得していても、その内容を深く理解しているとは限らない。大学では高校で学んだ内容を前提に授業を進めるのが自然であろう。特に数学を使用して授業を進める科目ではある程度の数学的な能力をもっているのが望ましい。本稿では数学基礎学力確認テストに対してデータ分析を行い、授業での活用を試みる。このような分析の試みは文献 [1, 2, 3] に見ることができる。

数学基礎学力確認テストは、高校の数学Ⅰと数学Aの内容から基本的な問題から応用まで幅広く出題している。全部で25問あり、数学Ⅰから19問、数学Aから6問である。回答はマークシートで、与えられた解答例から正解を選択する方法（5択）である。テストへのデータ分析方法には、現代テスト理論の一つである項目反応理論（item response theory, 以後IRTと略す）の適用を試みる。IRTはLordによって基本的なアイデアが与えられた理論で [4]、問題の難易度を考慮した成績評価が可能になる [5, 6, 7]。データ分析には統計解析ソフト「R」とパッケージ「irtosys」および「ICL」を使用した [7, 8, 9]。

また、得られた数学基礎学力確認テストの成績評価と数理授業科目である「数理基礎B」の受講者の試験成績と比較した。

まず、次節では古典的テスト理論による結果を述べる。つぎに第3節ではIRT分析を試みた結果を述べ、最後にまとめと今後の課題について述べる（第4節）

2. 古典的テスト理論

現在、基礎学力確認テストの成績評価は素点に基づいて行っている。テスト理論ではテストを構成する個々の問題を項目 (item) と呼ぶ。古典的テスト理論 (classical test theory, 以後 CTT と略す) では主に項目困難度、項目識別力や信頼性係数について調べられる。本節では、基礎学力確認テスト結果の項目困難度について述べるに留める。まず、受験者数は平成28年度入学生が121名、平成29年度入学生が168名で、その得点 (正答数) の分布を図1と図2に示す。各図において横軸が正答数、縦軸が割合である。

つぎに項目困難度は項目の正答率を使う。項目 (設問) ごとの正答率を平成28年度入学生の結果を図3に、平成29年度入学生の結果を図4に示す。

また、平成28年度と平成29年度の項目困難度の散布図を図5に示す。この相関係数は0.986で、受験者に対する項目困難度に関して入学年度に差はないとみており、図5から、困難度が0.3以下、つまり難しいレベルの項目が10問程度、困難度が0.8以上、つまり易しいレベルの項目が4問程度あることがわかる。

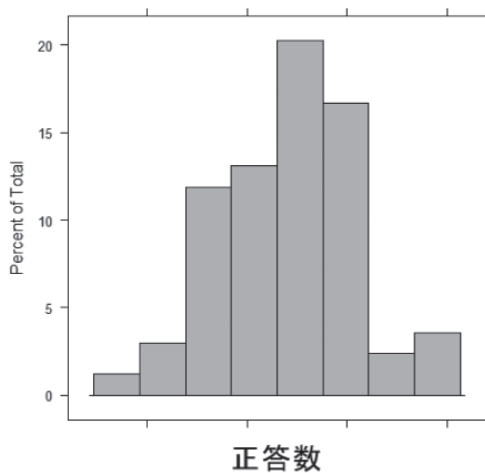


図1. ヒストグラム (平成28年度)

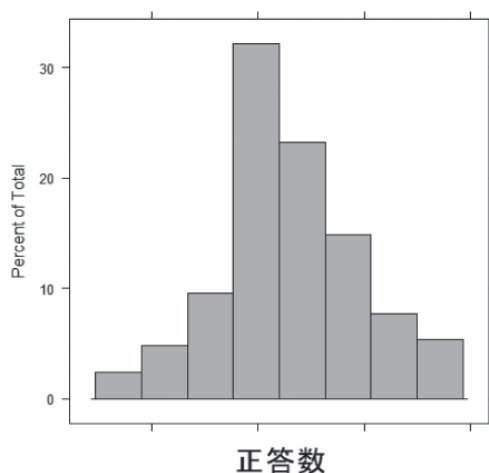


図2. ヒストグラム (平成29年度)

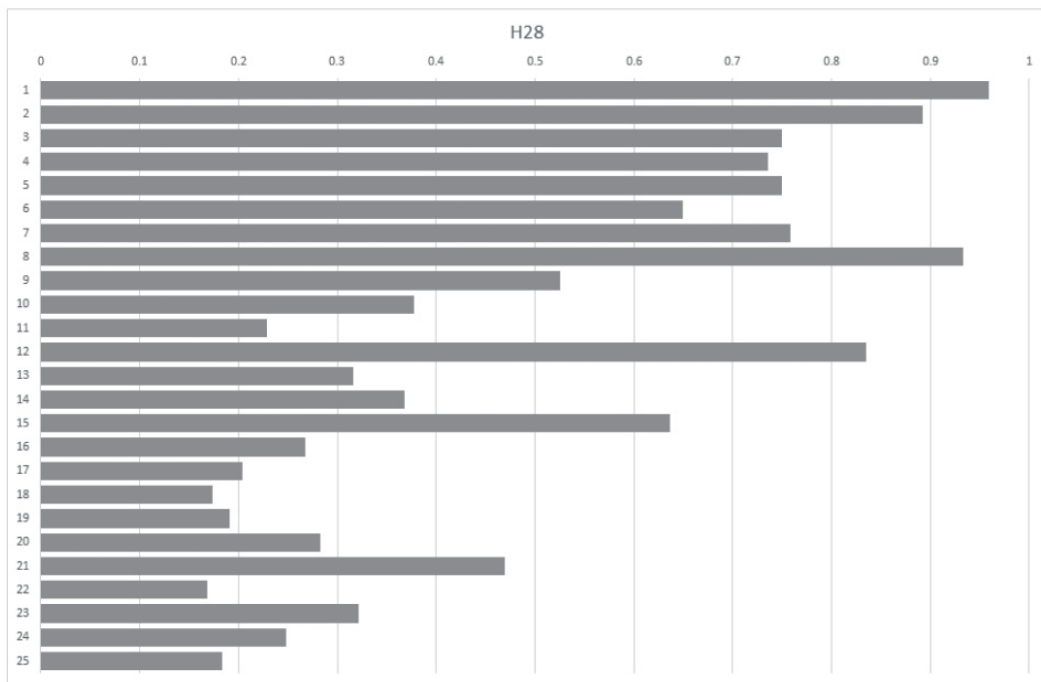


図 3. 正答率 (平成28年度)

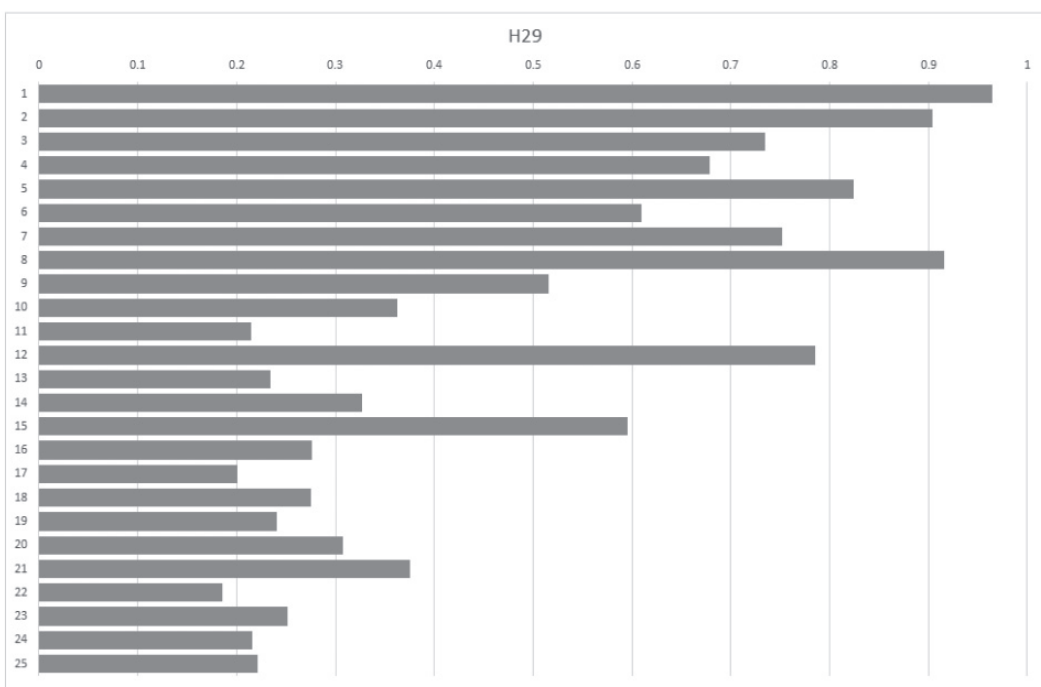


図 4. 正答率 (平成29年度)

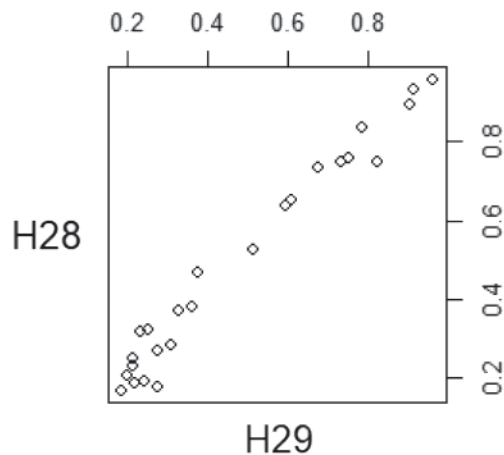


図5. 項目困難度 (CTT) の相関

3. IRT 分析の試み

IRT において、テスト項目の正答確率を項目の特徴を表す項目パラメタと受験者の能力パラメタ θ の関数で表現する。能力値 θ の受験者がある項目に正答する確率を θ の関数として表したものを項目特性関数 (item characteristic function, 以後 ICF と略す) と呼ぶ。

本稿では、ICF (ICC と呼ぶことが多い) を表現するモデルとして、つぎの 1 パラメタ・ロジスティックモデル (one parameter logistic model) を適用する。

$$P_j(\theta) = \frac{1}{1 + \exp(-(\theta - b_j))}$$

ここで、 b_j は項目困難度パラメタである。 b_j のパラメタ推定には最尤推定法を選択し、統計解析ソフト R とパッケージ「irtos」および「ICL」を使用し、求めた [7, 8, 9]。平成28年度入学生の項目ごとの結果を図6に、平成29年度入学生の結果を図7に示す。

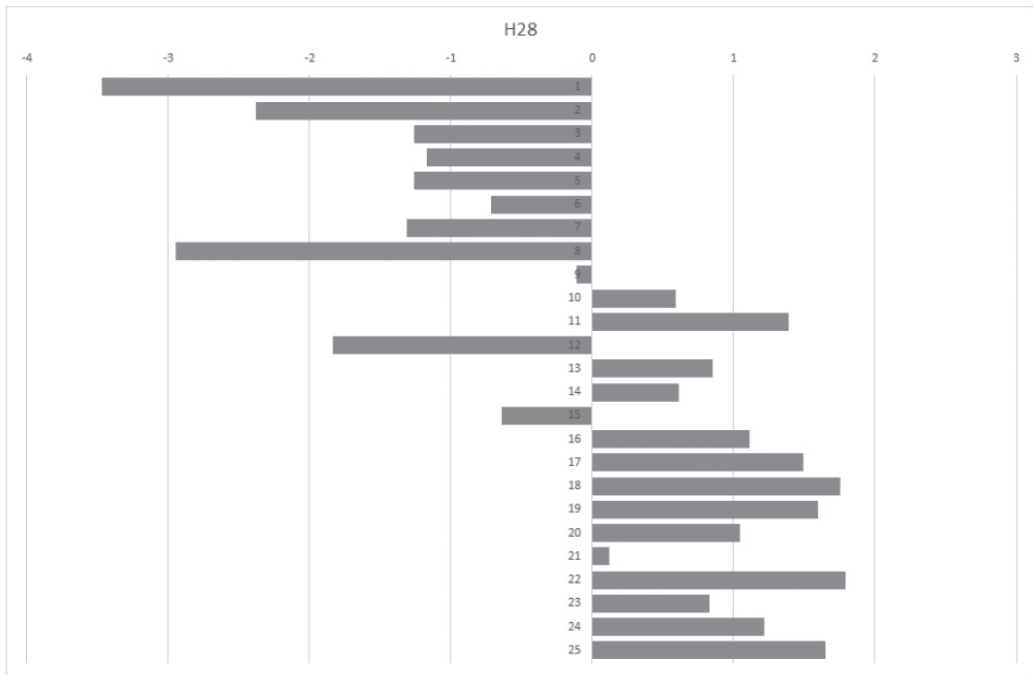


図6. 項目困難度 b_j (平成28年度)

これらの項目困難度は0辺りがふつうの問題、負の値が大きいほど易しい問題、正の値が大きいほど難しい問題と解釈する。また、平成28年度と平成29年度の散布図を図8に示す。ちなみに相関係数は0.983であった。

3.1. CTT との比較

CTTの項目困難度(正答率)とIRTの項目困難度との相関図を平成28年度入学生を図9に、平成29年度入学生を図10に示す。横軸がCTTの正答率、縦軸がIRTの b_j の値である。それぞれの相関係数は-0.990, -0.988なので、ほぼ直線上にあり、順に並んでいることがわかる。

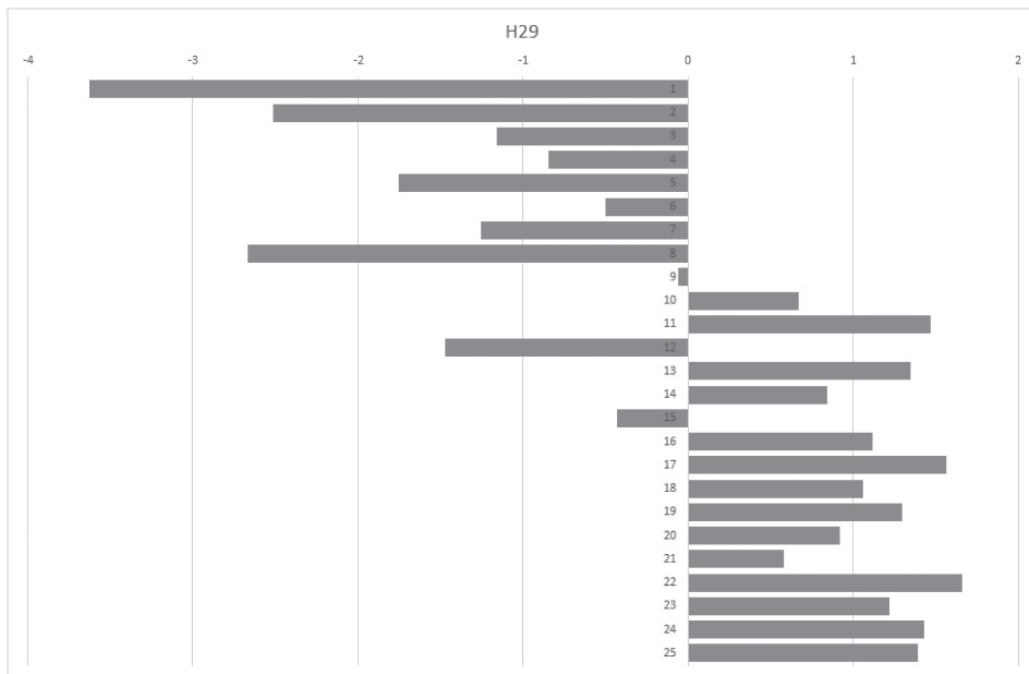


図7. 項目困難度 b_j (平成29年度)

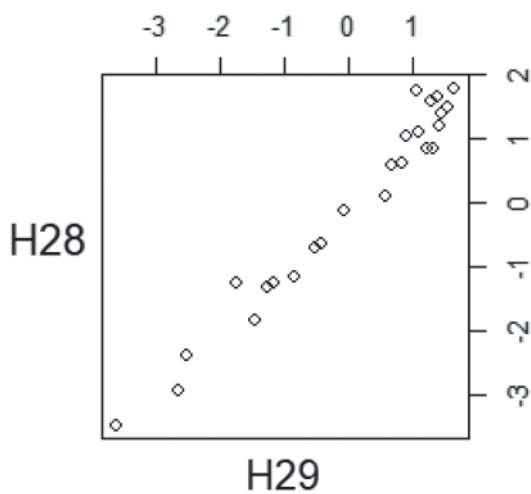


図8. 項目困難度 (IRT) の相関

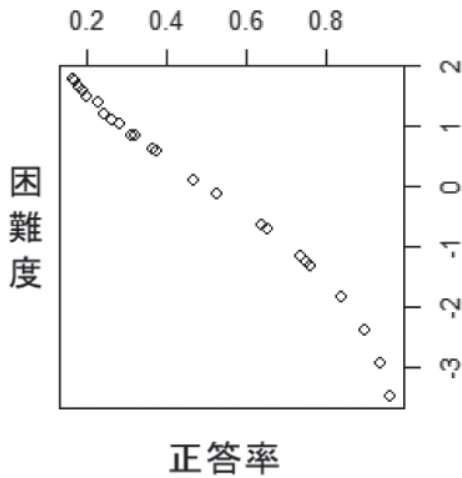


図9. CTT と IRT の相関 (平成28年度)

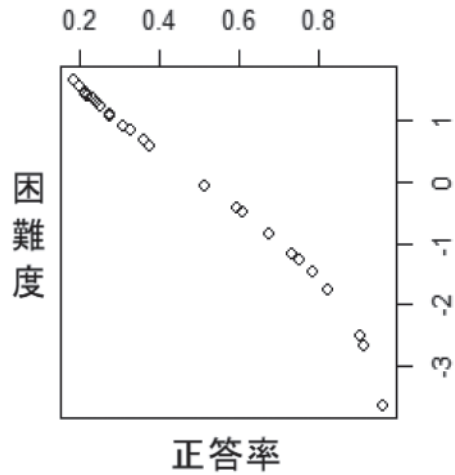


図10. CTT と IRT の相関 (平成29年度)

つぎに能力パラメタ θ の推定を行った。推定法には、項目困難度パラメタと同様に最尤推定法を選択し、パッケージを利用して結果を得た。能力値は項目困難度の値と同様、0辺りを平均と考え、正の方に大きくなるにしたがい能力が高く、負の方に大きくなるにしたがい能力が劣ると判断する。基礎学力確認テストの正答数と能力値の散布図を平成28年度入学生を図11に、平成29年度入学生を図12に示す。

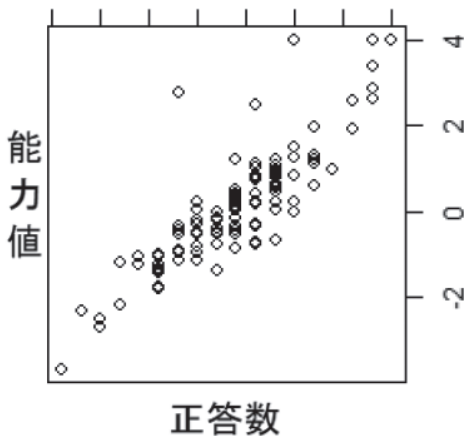


図11. 正答数と能力値の相関(平成28年度)

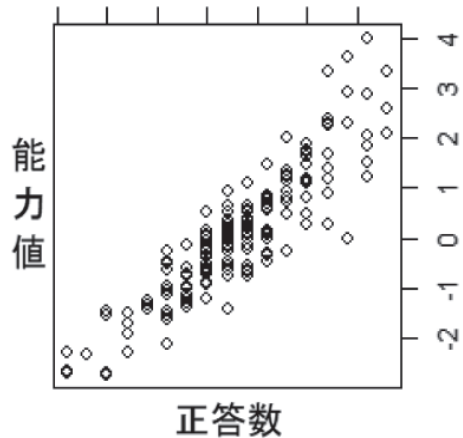


図12. 正答数と能力値の相関(平成29年度)

図11と図12を見ると、正答数が同じ場合でも、能力値にばらつきがあることに気づく。

3.2. 中間・期末試験評価との比較

平成29年度の授業科目「数理基礎B」で数学基礎学力確認テストを受けた受講者に対して中間・期末試験の成績との比較を行った。数理基礎Bの授業内容は高校数学の内容を土台に経済分野への応用ができることを目標にしている。中間試験と期末試験に分けたのは中間試験の内容が「数列とその応用」、期末試験の内容が「多項式の微分とその応用」と異なるためである。

「数理基礎B」の受講者の中で数学基礎学力確認テストを受験した学生は18名であった。この18名に対し、中間試験の結果(素点)と能力値との散布図を図13に、期末試験の結果(素点)と能力値との散布図を図14に示す。それぞれの相関係数は中間試験が0.747、期末試験が0.461となり、期末試験との相関が低いことがわかった。参考までに中間・期末の合計点との散布図を図15に、中間と期末の散布図を図16に示す。図16をみると、中間と期末の相関係数が0.450と相関が低い。相関が低い要因は、中間・期末試験の内容が数学基礎学力確認テストにない問題であること、中間が数学B、期末が数学IIと試験の内容が異なること、高校で数学IIや数学Bの内容を教わっていない学生がいること、試験では応用力をみる問題が含まれていること、などが考えられる。

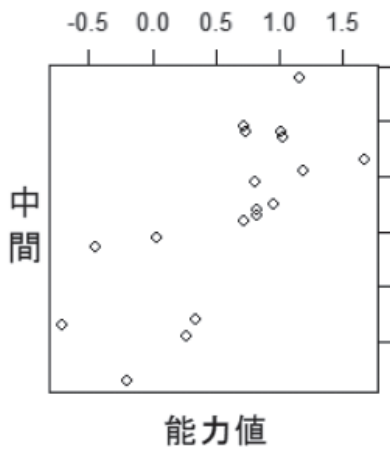


図 13. 中間試験との相関

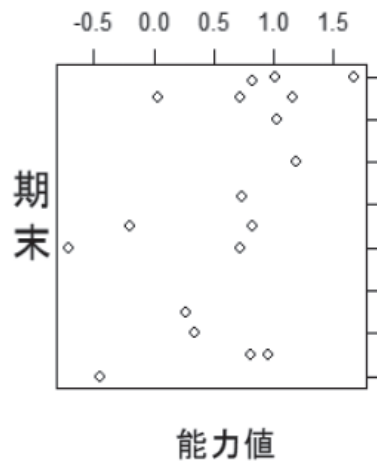


図 14. 期末試験との相関

4. まとめと今後の課題

本稿では、入学直後に行われる基礎学力確認テスト結果に項目反応理論 (IRT) の適用を試み、その分析結果と授業の中間・期末試験の成績との比較を行った。IRTのモデルとして1パラメタ・ロジスティックモデルを適用した。テスト形式が5択であることを考慮すると、当て推量パラメタのある3パラメタ・ロジスティックモデルの適用が最適である。

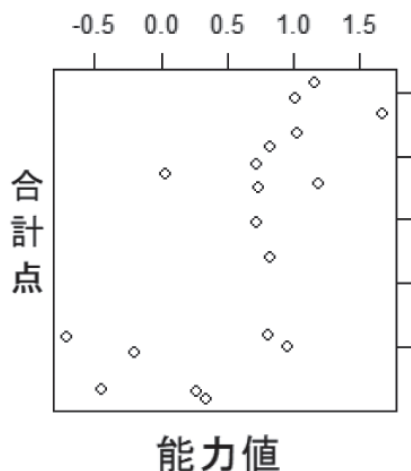


図 15. 中間+期末との相関

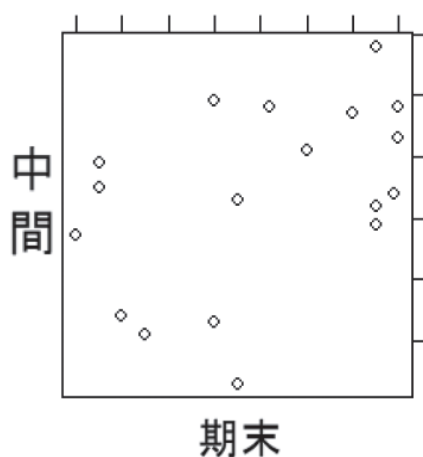


図 16. 中間と期末の相関

しかし、実際によく適用される2パラメタ・ロジスティックモデルに適用を試みると、項目パラメタの推定不能に遭遇する。原因として

- 標本数が少ない
- テスト項目の設問の質
- 受験者の学力のばらつき

などが考えられる。今後は古典的テスト理論（CTT）で使われる項目識別力やトレースラインによる図示化により、データに前処理を施すなど詳細な分析を行うことで、IRTの適用を試みる。図11と図12で正答数が同じでも能力値にばらつきがあることがわかった。能力値は推定であるから、標本数が少ないこともあり、さらなる分析を進めることでより精度を高めたい。また、IRTで使用される推定法は数値計算技術によるところが大きい。本稿ではパッケージを利用して推定値を求めたが、実際のデータに直接数値計算法の適用を試み、推定値を求めることで、計算技術の改良を行いたい。

他方 IRT 分析結果の授業への活用があげられる。多様で学力の幅広い学生をどの程度の能力ならば想定した授業についていけるのかは教員として授業運営の課題といえる。さらなるデータ分析と授業評価を教材作成にフィードバックすることで、学生の能力向上が今後の課題である。

謝辞

本研究は平成28年度岐阜聖徳学園大学研究助成金を受けて行われました。

参考文献

- [1] 巽靖昭、東晋司、児玉俊介、佐藤崇、澤口隆、2012、「ミクロ・マクロ経済学演習科目の教育効果に関する実証研究」、京都大学高等教育研究、18、pp 11-23
- [2] 廣瀬英雄、2016、「数学プレースメントテスト成績と期末試験成績の比較」、広島工業大学紀要教育編、15、pp 35-38
- [3] 廣瀬英雄、2016、「異なる集団間での項目反応理論パラメータ —数学プレースメントテスト成績の大学間比較の一例」、広島工業大学紀要教育編、15、pp 39-43
- [4] Lord, F. M., 1952, “A theory of test scores”, Psychometric Monograph, No7
- [5] 村木英治、2011、『項目反応理論』、朝倉書店、東京
- [6] 植野真臣、莊島宏二郎、2010、『学習評価の新潮流』、朝倉書店、東京
- [7] 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、東京
- [8] Partchev, I., 2014, “irtoys: Simple interface to the estimation and plotting of IRT models”, Package Version 0.2.1 , <https://CRAN.R-project.org/package=irtoys>, 2018/04/27 アクセス
- [9] Hanson, B. P., 2002, “IRT Command Language”, <https://sourceforge.net/projects/ssm/>, 2018/04/27 アクセス