

## 項目反応理論

野村 義明<sup>1)</sup>, 松山 知明<sup>2)</sup>, 花村 裕之<sup>2)</sup>, 深井 穂博<sup>3)</sup>

### Application of item response theory for dental research

Yoshiaki Nomura<sup>1)</sup>, Tomoaki Matsuyama<sup>2)</sup>, Hiroyuki Hanamura<sup>2)</sup>, Kakuhiro Fukai<sup>3)</sup>

<sup>1)</sup> 鶴見大学歯学部探索歯学講座, <sup>2)</sup> 川崎市歯科医師会, <sup>3)</sup> 深井保健科学研究所

キーワード：項目反応理論、疫学指標、質問紙

#### 要 旨

古典テスト理論で用いられる識別係数では、予め設定した合計点を用いるため、個々の項目に関するウェイトを算出することができない。項目反応理論は個々の質問項目のウェイトと（能力：Ability）合計点を同時に推定する方法である。歯科ではDMF、歯周病のインデックス等、合計点を使用しているものが多いため、項目反応理論を適用できる場合が多い。本稿では、項目反応理論の基本的な考え方とその適用例を示した。

#### 合計に対する構成要素による分析

歯科関連の指標のなかには各項目の合計点で評価する指標が多々存在する。DMFはD歯数、M歯数、F歯数の合計で評価する。このような指標は他にも、多くの歯周病の疫学指標をはじめ、質問紙では、日本歯科医師会による標準的な質問票などがある。

DMFの分析を行う場合、回帰分析により

$$\text{DMF} = a_1 \times \text{D歯数} + a_2 \times \text{M歯数} + a_3 \times \text{F歯数} + \beta$$

のような合計を従属変数、構成要素を独立変数とするような回帰式を用いることはない。

能力の判定にはテストを用いることが多々ある。テストは個々問題の正答から合計点を算出して、能力を評価する。そのためには、個々の問題が適切であるかを評価し、適切なテストを構成する必要がある。この場合、合計点と各問題の関連を分析し、適切な指標で評価する必要がある。合計点と各問題の関連の分析手法は、テスト理論と呼ばれる理論体系が構築されている。

#### テスト理論

テスト理論は大きく分けて、古典テスト理論と項目反応理論（Item Response Theory；Item

#### 【著者連絡先】

〒230-8501 神奈川県横浜市鶴見区鶴見2-1-3

鶴見大学歯学部探索歯学講座

野村義明

TEL：045-580-8462 FAX：045-573-9599

E-mail：nomura-y@tsurumi-u.ac.jp

受付日：2018年11月30日 受理日：2018年12月14日

Latent Theory : IRT) を中心とした現代テスト理論がある<sup>1)</sup>。古典テスト理論とは、観測得点 = 真の得点 + 測定誤差と考え、それぞれの分散を用いて観測得点に真の特定が反映されている程度を評価し、テストの信頼性を評価しようとする理論である。クロンバックの  $\alpha$  係数は内的整合性という各質問項目 (変数) が全体として同じ概念や対象を測定したかどうかを評価する指標で、古典テスト理論の代表的な指標として広く用いられている。合計得点が高い者の正答率が高くなっているか、合計得点が高い者の正答率が低くなっているかを評価する識別指数も古典テスト理論の代表的な指標である。古典テスト理論では、合計点を固定して分析を行うが、IRTでは、合計点を各受験者の特性の大きさを表す実数値である人パラメータとして扱う。この人パラメータと各項目の被験者の能力を識別する力を表す実数値とする識別パラメータとの関連を分析するもので、被験者の能力である標本依存性と項目の難易度であるテスト依存性にとらわれずに不変的に受験者の能力値とテスト項目の難易度を求められるとされている。つまり、サンプルに依存しない分析が可能であるとされている。

### IRT

IRTでは、ロジステックモデル、プロビットモデルがあるが、ロジステックモデルが一般的に用いられているため、ロジステックモデルについて解説する。人パラメータ ( $\theta$ ) と識別パラメータのみを用いるものを1パラメータロジステックモデル、困難度パラメータを加えたものを2パラメータロジステックモデル、さらに当て推量パラメータを加えたものを3パラメータロジステックモデルと呼んでいる。

ロジステックモデルでは上記の人パラメータ、識別パラメータの他、難易度 (困難度) パラメータ、当て推量パラメータを導入して分析を行うことができる。人パラメータ ( $\theta$ ) と識別パラメータ ( $a_i$ ) のみを用いるものを1パラメータロジステックモデル (1PL)、困難度パラメータ ( $b_i$ ) を

加えたものを2パラメータロジステックモデル (2PL)、さらに当て推量パラメータ ( $c_i$ ) を加えたものを3パラメータロジステックモデル (3PL) と呼んでいる。それぞれの式は以下のようになっている。

1パラメータロジステックモデル

$$p_i(\theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

2パラメータロジステックモデル

$$p_i(\theta) = \frac{1}{1 + e^{-D a_i (\theta - b_i)}}$$

3パラメータロジステックモデル

$$p_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-D a_i (\theta - b_i)}}$$

項目特性曲線 (ICC : item characteristic curve) と項目情報曲線 (IFC : Item Information Curve) が得られる。1PLではICCの傾きが一定であるのに対して、2PLでは傾きが項目によって異なる。1PL、2PLでは曲線の立ち上がりがほぼ0であるのに対して、3PLでは曲線の立ち上がりが項目によって異なるという特徴があり、1PLよりも2PL、2PLよりも3PLの方が項目の特性を判別しやすい。

### 項目反応曲線 (ICC) と項目情報曲線の解釈

図1に日本歯科医師会による標準的な質問票の3PLによる分析結果を示す。項目は一部抜粋している。図の横軸はIRTでは能力を示し、-3から3の間で0.01以上異ならないようにロジスティック関数を累積正規分布関数に近似している。日本歯科医師会による標準的な質問票では、「はい」と回答された項目の多さを示している。Qは7典型的なシグモイドカーブを示している。これと比較してQ12左にシフトしており、Q7と比較して「はい」と回答した項目数が少ない人で、「はい」と答えた数を判別するのに有用な項目である。Q3はQ7と比較して曲線が右にシフトしており、曲線の立ち上がりも急である。テストでは一定水準以下の能力では解答することが難しく、能力が一

定水準を超えると確実に解答できる良問に相当する。これらの項目は項目情報量が多い。Q5は困難度が高い項目である。テストでは最も能力の高い者の集団でも全員が正解を得られないような項目である。これらと比較して、Q13では横軸が-4でも高いを示し、曲線の傾きも緩やかである。このような曲線は「はい」と回答する数に対する判別力は小さい。Q9は能力が低い者の方が「はい」と回答した者が多いという望ましくない質問項目である。これらの項目は「はい」と回答した数を判別できる項目情報量が少ないことがわかる。

同じ項目を1PL、2PL、3PLで分析した項目反応曲線の比較を図2に示す。1PLと比較して2PLの方が、2PLと比較して3PLの方が各項目の特性を把握しやすいのがわかる。

### 日本歯科医師会標準的な質問票について

日本歯科医師会標準的な質問票はPrecede Proceed modelをベースにして作成された、保健指導に有用な質問票である。今回の分析結果では、幾つかの項目で適切でない可能性がある項目がみられた。データをベースにした分析結果を活用し、さらに改良の余地があることが示唆された<sup>2)</sup>。

### IRTのその他の応用例

乳歯う蝕の有無を1PLで分析した結果で一部抜粋した結果を図3に示す。上顎のDが最もう蝕に罹患しやすく、続いて、上顎のCがう蝕に罹患しやすく、下顎のBが最もう蝕に罹患しにくいことがわかる。

山本式総義歯咀嚼能率判定表の分析結果を図4に示す。ごはんは咀嚼能力が低い者でもかむこと

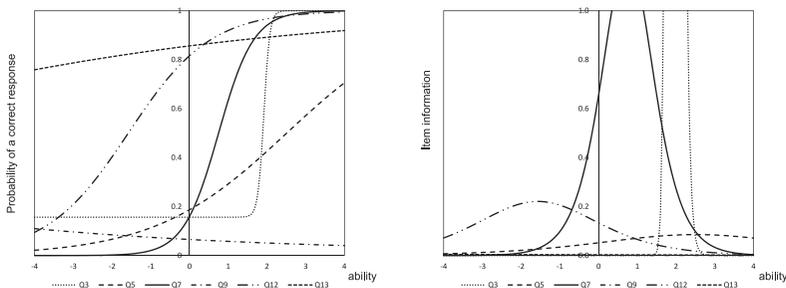


図1 典型的な項目反応曲線と項目情報曲線

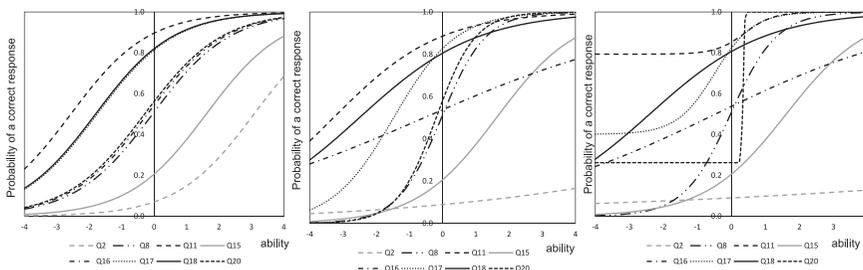


図2 1PL、2PL、3PLの項目反応曲線の比較

同じ項目を1PL、2PL、3PLで分析した結果  
1PLと比較して2PLの方が、2PLと比較して3PLの方が各項目の特性を把握しやすいのがわかる。

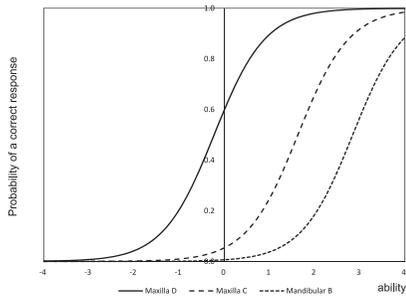


図3 乳菌う蝕の罹患に対する項反応曲線

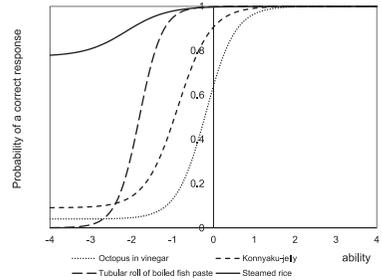


図4 山本式総義菌咀嚼能率判定表による項目反応曲線

ができる。こんにゃく、ちくわ、たこの酢漬けの順に噛みにくいことがわかる。

#### その他IRTについて

IRT分析に必要なサンプル数は1PLで100例、2PLで200例、3PLで300例と言われているが、分析する項目数に依存すると思われる。今回の分析は変数をすべて2値化して行ったものである。Likert scaleのような2値でないものでも Samejimaの Graded response modelsで対応することができる<sup>3)</sup>。IRTはサンプルに依存しない分析が可能であるとされているが、項目の特性とともに能力を推定できることが特徴で、実際に分析を行うと、結果はサンプルに依存してしまう。汎用の統計ソフトであるSPSS等にはIRTは実装されていない。現時点では専用のソフトが販売されているため、それらを利用するか、SAS、R等でプログラムを組んでゆく必要がある。また、IRTに関する優れた日本語の解説書は多々あるためそれらを参考にして下さい<sup>3-7)</sup>。

#### 文献

- 1) 池田 央 現代テスト理論 (行動計量学シリーズ) 朝倉書店 (1994/10)
- 2) Nomura Y et al. Precede-Proceed model based questionnaire and saliva tests for oral health checkup in adult Journal of Oral Science (in press)
- 3) Fumiko Samejima Evaluation of mathematical models for ordered polychotomous responses. Behaviormetrika 23巻 (1996) 1号17-35.
- 4) 加藤健太郎, 山田剛史, 川端一光 Rによる項目反応理論 オーム社 (2014/7/26)
- 5) 豊田秀樹 項目反応理論 [入門編] (第2版) (統計ライブラリー) 朝倉書店; (2012/8/24)
- 6) 豊田秀樹 項目反応理論 [中級編] (統計ライブラリー) 2013/10/23
- 7) 豊田秀樹 項目反応理論 事例編 - 新しい心理テストの構成法 (統計ライブラリー) 2002/7/1
- 8) 豊田秀樹 項目反応理論・理論編 - テストの数理 (統計ライブラリー) 単行本 - 2005/6/1