項目間の関連を分類するための指標の研究

A research for the marker in classifying item relation

浅尾 彰俊
Akitoshi ASAO
元高校教員
Former high school teacher

＜あらまし＞　項目間の関連を示す指標として相関係数がよく利用されるが，相関係数は対称な指標となっており，原因・結果の関係を表すものではない．これに対して項目関連構造分析において用いられる順序性係数は,「学習者から見て一方の項目ができなければ他方の項目はできない」という形で非対称な順序関係を判別するのに用いられる．この研究においては，まず回帰係数，相関係数，順序性係数の関係を数式変形として調べる．次に順序性係数を用いて項目間の関係を判別するための分類表を考えるとともに，データが正答率である場合にも順序性係数を拡張する方法を考える．また，これらが実際のデータに適合するかどうかを調べる．

＜キーワード＞　データ解析，数理モデル，回帰係数，順序性係数，項目関連構造分析

1. はじめに

中学校の教科書で教材作成者から見た内容系統表は教科書会社から公開されていることがある．これに対して学習者の理解状況を示す単元間の関連はＳＰ分析，項目関連構造分析などによって研究されることが多い．

当研究では，項目関連構造分析で用いられる順序性係数を手掛かりとして項目間の関連を分類する方法を考え，実データへの適合性を調べる．

2. ２×２分割表と回帰係数の関係

テスト結果を表す正誤情報は 1-0 データで表され，２項目からなるテストの正誤情報を集計すると次のような2×2分割表が得られる．表 1 における文字は各々観測度数を表し，a（２項目とも正答）, b（ j 正答，ｋ誤答）, c（ j 誤答，ｋ正答）, d（２項目とも誤答）, N（合計）とする．

表 1

	項目 k		小計
項目 j	a	b	a+b
	c	d	c+d
小計	a+c	b+d	N

正答率が p である項目 j によって正答率が q である項目ｋを回帰分析して

$$q=\alpha+\beta p \cdots(1)$$

の関係が得られたとき，回帰式の定数項は j が誤答であるときにｋが正答となる条件付き確率を表し，回帰係数は j が正答であるときにｋが正答となる条件付き確率と j が誤答のときにｋが正答となる条件付き確率の差を表していると考えられる．

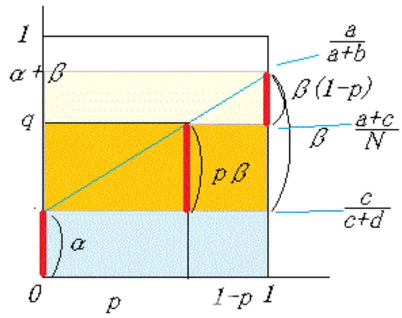


図 1

3. 順序性係数と回帰係数の関係

項目関連構造分析に用いられる順序性係数$\gamma^*{}_{jk}$は表１の観測度数 a，b，c，d を用いて

$$\gamma^*{}_{jk}=1-\frac{cN}{(c+d)(a+c)}\cdots(2)$$

と定義され，$\gamma^*{}_{jk}$があるしきい値（通常は 0.5）以上のとき，ｊ→ｋの順序性があるとされ「項目ｊができないならば項目ｋはできない」と解釈されている．

この式は

$$\gamma^*{}_{jk}=1-\frac{\dfrac{c}{c+d}}{\dfrac{a+c}{N}}\cdots(3)$$

と変形できるから，順序性係数は(1)の定数項α，回帰係数βを用いて表すことができる．

$$\gamma_{jk}^{*} = 1 - \frac{\alpha}{q} = \frac{\beta p}{q} \cdots (4)$$

4. 順序性係数と相関係数の関係

表１の分割表から求められる順序性係数と相関係数ｒは次の関係を満たす．

$$r = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \cdots (5)$$

$$\gamma^{*}{}_{jk} = \frac{ad - bc}{(a+c)(c+d)} \cdots (6)$$

したがって，順序性係数は相関係数を用いて表すこともできる．

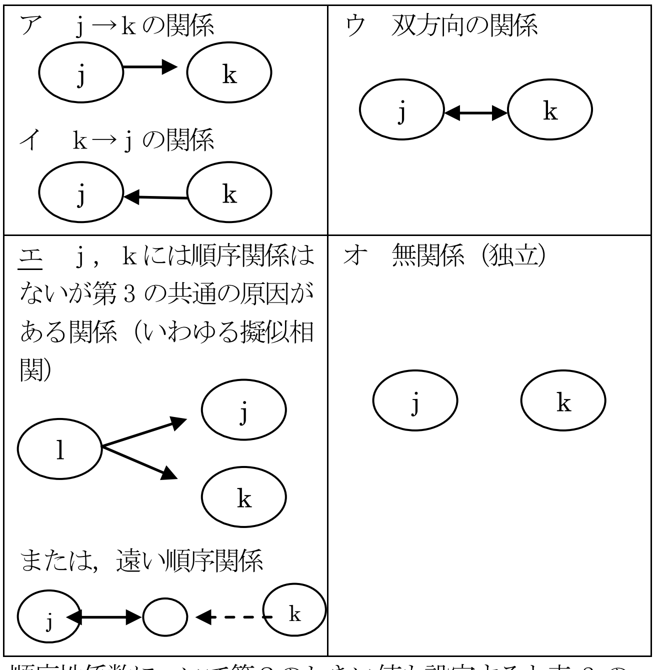
$$\gamma^{*}{}_{jk} = \sqrt{\frac{p(1-q)}{q(1-p)}}\, r \cdots (7)$$

なお，順序性係数は(2)または(4)によって求められるので，実際の数値を求める上では(7)はなくてもよいが，当研究においては(7)を根拠として，相関係数の有意性の判断によって順序性係数の有意性の判断を行った．

5. 順序性係数による項目関連の分類

順序性係数がしきい値以上のとき２つの項目間には順序性があると判断し，項目ｊは項目ｋに優先的に教えるべきであると考えるのが通常の解釈であるが，ここでは，次の図のような２項目間の関係を区別するための指標という側面に絞って考える．

表２



順序性係数について第２のしきい値も設定すると表２の関係は次の表３の分類表のいずれかに当てはまる．

以下においては，実際のデータに適用したときの経験則として C_1=0.3，C_2=0.5 を用いた．また，擬似相関は観測項目間の関係を確認することのみに用い，潜在要因には触れなかった．

なお，このような関係は因果関係と呼ばれることもあるが，当研究では順序関係という用語を用いる．

表３

k→j / j→k	γ_{kj}^{*}<C_1	$C_1 \leqq \gamma_{kj}^{*}$<$C_2$	$C_2 \leqq \gamma_{kj}^{*}$
γ_{jk}^{*}<C_1	オ	エ	イ ←
$C_1 \leqq \gamma_{jk}^{*}$<$C_2$		…	
$C_2 \leqq \gamma_{jk}^{*}$	ア →		ウ ⟷

6. 正答率データへの順序性係数の拡張

(4)を元にデータが正答率である場合に順序性係数を拡張する方法を考える．

小数値のまま(4)を適用する場合は次の式によって計算した．ここで S_{jk} は共分散，S_j は標準偏差を表す．

$$\beta = \frac{s_k}{s_j} r \cdots (8)$$

$$\gamma^{*}{}_{jk} = \frac{\beta p}{q} = \frac{p\, s_{jk}}{q\, s_j^{2}} \cdots (9)$$

このとき相関係数との関係(7)も書き換える．

四捨五入のように何らかのボーダーラインを設定して1-0 データに変換するときは順序性係数の元の定義を用いた．

7.　実データに適用したときの適合性

実際に答案を集計して，5.の分類表及び 6.の拡張の適合性を調べたところ，次の結果が得られた．

(1) データが1-0の正誤情報で与えられる場合については，5.の分類表により問題なく分類できるが，独自提案の分類エを使える場面は少なく，結果の確認に留まる．

(2) データが正答率で与えられた場合については，1-0 情報に変換する方法よりも正答率のまま順序性係数を求める方が項目間の関連をよく表していると考えられる．

8. まとめと今後の課題

ここで用いた関係式や数式の解釈には他の文献で確認できない試案(_ 印)も含まれているが，実データに適用した結果は悪くない．

当研究で扱った方法は，観測項目間の関連分析に限られるが難しい数学理論や特別なソフトを要しないため Excel 上で簡単に行うことができる．（マクロは使う）また，百分率からなる数十項目の統計資料について，項目間の関連を探索的に分析するのに適していると考えられる．

当研究では，データが正答率で与えられる場合に実データへの適合性を内容に踏み込んで判断したが，何らかの客観的な基準を併用する方がよいと考えられる．

参考文献

日本教育工学会編 (2000)「教育工学事典」．実教出版

竹谷誠 (1980) IRS テスト構造グラフの構成法と活用法

Chevalaz, Gerard M.; Tatsuoka, Kikumi K. (1983) A Comparative Analysis of Two order Analytic Techniques

小島隆矢 (2003)「Excel で学ぶ共分散構造分析とグラフィカルモデリング」. オーム社