

# 項目間の関連を分類するための指標の研究

元高校教員 浅尾 彰俊

二項目間の関連の強さを示す指標として相関係数がよく利用される。ただし相関係数は二つの項目に関して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 (IRS) 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」という順序関係を判別するのに用いられており非対称な指標となっている。当研究においては次の内容について述べる。

1. ~2. 相関係数, 順序性係数, 回帰係数の関係について筆者の解釈を示し, 回帰係数から順序性係数を求める式を考える。
3. 順序性係数から項目間の関係を判別するための分類表を試作する。
4. 順序性係数が満たすべき性質を確認する。
5. 1-0 データからなる答案を集計して, 5. の分類表で項目関連構造がどの程度解明できるかを調べる。
6. 3. において得られた分類表を各項目が正答率となっている場合に拡張する方法を考える。
7. 正答率データからなる答案を集計して, 6. で述べた方法によって項目関連構造がどの程度解明できるかを調べる。

## 1. 回帰係数と $2 \times 2$ 分割表の関係

次の表 1 のように 2 つの項目  $j$  と  $k$  について何件かの正誤情報が与えられているとする。データは列方向に表示されるのが通例であるが紙面の都合で行方向に表示する。また, 1. から 5. では正誤情報 (1-0 データ) のみを扱う。

表 1

項目 $j$	1	1	1	1	1	1	1	0	0	0	0	0	0	
項目 $k$	1	1	1	1	1	0	0	0	1	1	1	1	0	0
集計	a=5				b=3			c=4			d=2			

表 1 のような 1-0 データが  $N$  件 (表 1 では 14 件となっている) あるとき, これらを  $2 \times 2$  分割表に集計したものを表 2 とする。

項目  $j$  の 1-0 データを変数  $X_{ji}$  ( $i=1 \sim N$ ), 項目  $k$  の 1-0 データを変数  $X_{ki}$  ( $i=1 \sim N$ ) で表す。

項目  $j$  の正答率を  $p$ , 項目  $k$  の正答率を  $q$  で表すとき,

$$p = \frac{1}{N} \sum_{i=1}^N X_{ji} = \frac{a+b}{N} \cdots (1)$$

$$q = \frac{1}{N} \sum_{i=1}^N X_{ki} = \frac{a+c}{N} \cdots (2)$$

項目  $j$ ,  $k$  について説明変数を  $X_j$ , 目的変数を  $X_k$  とする回帰分析を行って,

$$X_{ki} \doteq \alpha + \beta X_{ji} \cdots (3) \quad (\alpha \text{ は定数項, } \beta \text{ は回帰係数})$$

という回帰式が得られるとき,  $\alpha$ ,  $\beta$  と  $a$ ,  $b$ ,  $c$ ,  $d$  には次の関係がある。

$$\alpha = \frac{c}{c+d} \cdots (4)$$

項目  $j$  が誤答であるときに項目  $k$  が正答となる確率を表す。

$$\beta = \frac{a}{a+b} - \frac{c}{c+d} \cdots (5)$$

「項目  $j$  が正答であるときに項目  $k$  が正答となる確率」と「項目  $j$  が誤答であるときに項目  $k$  が正答となる確率」の差を表す。

表 2

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

(4)(5)は次のように示せる.

(3)より

$$\sum_{i=1}^N X_{ki} = \sum_{i=1}^N (\alpha + \beta X_{ji})$$

$$\frac{1}{N} \sum_{i=1}^N X_{ki} = \frac{1}{N} \sum_{i=1}^N (\alpha + \beta X_{ji})$$

だから  $q = \alpha + \beta p \cdots (6)$

ここで  $p=0$  のとき  $q=\alpha$  だから,  $\alpha$  は項目  $j$  が誤答であるときの項目  $k$  の正答率を表す.

よって  $\alpha = \frac{c}{c+d}$  となる.  $\rightarrow(4)$

$p=1$  のとき  $q=\alpha+\beta$  だから,  $\alpha+\beta$  は項目  $j$  が正答であるときの項目  $k$  の正答率を表す.

$$\alpha + \beta = \frac{a}{a+b}, \text{ したがって } \beta = \frac{a}{a+b} - \frac{c}{c+d} \text{ となる. } \rightarrow(5)$$

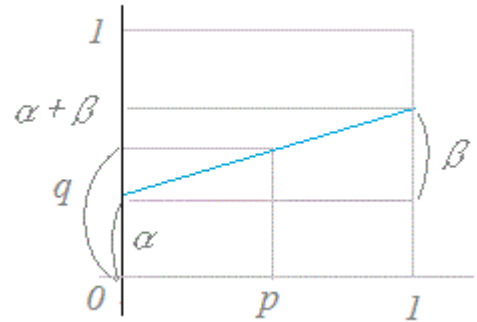


図1

## 2. 相関係数と順序性係数の関係

2.1 項目  $k, j$  間のピアソンの積率相関係数 (以下, 単に相関係数という) は次のように定義される.

$$\text{相関係数} = \frac{\text{項目 } j \text{ と項目 } k \text{ の共分散}}{\sqrt{\text{項目 } j \text{ の分散}} \sqrt{\text{項目 } k \text{ の分散}}} = \frac{1}{N} \frac{\sum (x_{ji} - m_j)(y_{ki} - m_k)}{s_j s_k}$$

ここで,  $x_i, y_i$  は項目  $j, k$  の変数の値,  $m_j, m_k$  は項目  $j, k$  の平均,  $s_j, s_k$  は項目  $j, k$  の標準偏差,  $N$  はデータ件数を表わす.

項目  $j, k$  が 1-0 データであるときは平均や標準偏差は簡単となり, これらを  $p, q$  で表わすと,

$$p = \frac{a+b}{N}, \quad q = \frac{a+c}{N}, \quad m_j = p, \quad m_k = q$$

$$s_j = \sqrt{p(1-p)}, \quad s_k = \sqrt{q(1-q)}$$

が成り立つ.

これらを用いて相関係数  $r$  は次のように変形できる.

$$r = \frac{1}{\sqrt{p(1-p)} \sqrt{q(1-q)}} \left( \frac{a}{N} - pq \right) \quad \dots(7)$$

$$r = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad \dots(8)$$

2.2 項目関連構造分析に用いられる順序性係数  $\gamma^*_{jk}$  は表 2 の  $2 \times 2$  分割表の観測度数  $a, b, c, d$  を用いて

$$\gamma^*_{jk} = 1 - \frac{cN}{(c+d)(a+c)} \quad \dots(9)$$

と定義され,  $\gamma^*_{jk}$  がしきい値 (通常は 0.5) 以上のとき, 通常は「項目  $j$  ができないならば項目  $k$  はできない」と解釈され,  $j \rightarrow k$  の順序性があるとされる. この解釈においては, 表 2 における観測度数  $c$  が小さいことが目安となっている.

$$\gamma_{jk}^* = 1 - \frac{cN}{(c+d)(a+c)} = \frac{(c+d)(a+c) - cN}{(c+d)(a+c)} = \frac{ad - bc}{(c+d)(a+c)} \dots(10)$$

と変形できるから相関係数  $r$  は順序性係数  $\gamma_{jk}^* = \frac{ad - bc}{(a+c)(c+d)}$  を用いて

$$r = \sqrt{\frac{(a+c)(c+d)}{(a+b)(b+d)}} \gamma_{jk}^* = \sqrt{\frac{q(1-p)}{p(1-q)}} \gamma_{jk}^* \dots(11)$$

と表すことができる。

これを逆に解けば、 $\gamma_{jk}^* = \sqrt{\frac{p(1-q)}{q(1-p)}} r$  となる。

また前節で述べた回帰係数  $\beta$  を用いると

$$\beta = \frac{a}{a+b} - \frac{c}{c+d} = \frac{ad - bc}{(a+b)(c+d)}$$

だから

$$\frac{p\beta}{q} = \frac{ad - bc}{(a+b)(c+d)} \frac{a+b}{N} \frac{N}{a+c} = \frac{ad - bc}{(a+b)(c+d)} = \gamma_{jk}^*$$

以上 1.~2.の内容は次のようにまとめることができる。

順序性係数  $\gamma_{jk}^*$  は、回帰係数  $\beta$  と項目  $j$  の正答率  $p$ 、項目  $k$  の正答率  $q$  を用いて表すことができる。

$$\gamma_{jk}^* = \frac{\beta p}{q} \dots(15)$$

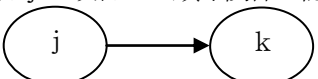
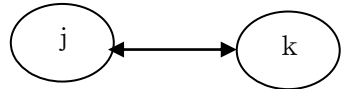
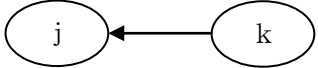
順序性係数は相関係数と正答率を用いて表すことができる。

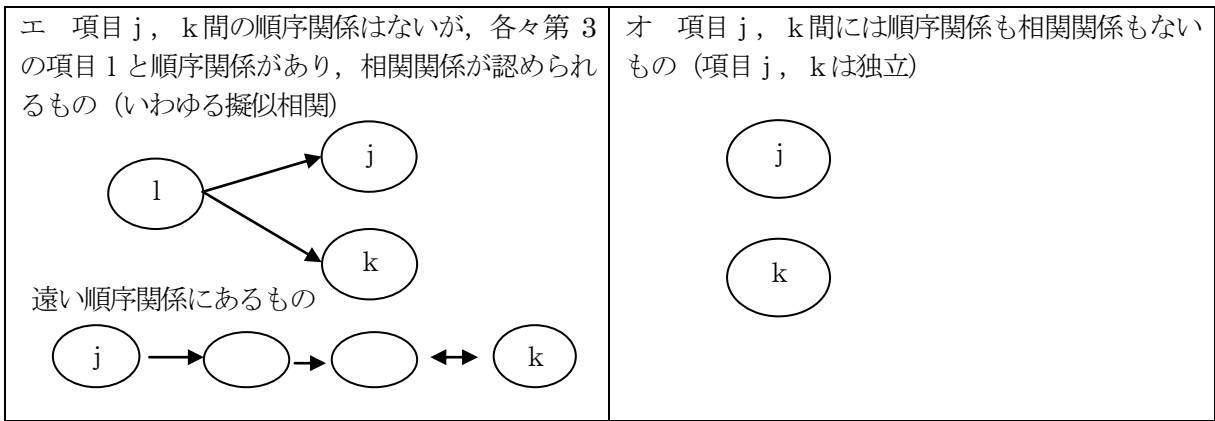
$$\gamma_{jk}^* = \sqrt{\frac{p(1-q)}{q(1-p)}} r \dots(16)$$

### 3. 順序性係数を用いた項目関係の分類

項目間の関連については次の表 3 のような関係が考えられる。項目関連構造分析において用いられる矢印は数学記号における必要条件、十分条件と逆になっており、「 $j$  ができないならば  $k$  はできない」という関係が  $j \rightarrow k$  で表されている。このような関係は因果関係と呼ばれることがあるが、以下においては順序関係と呼ぶ。

表 3

<p>ア 項目 <math>j \rightarrow</math> 項目 <math>k</math> の順序関係が認められるもの</p> 	<p>ウ 項目 <math>j, k</math> 間に双方向順序関係が認められるもの</p> 
<p>イ 項目 <math>k \rightarrow</math> 項目 <math>j</math> の順序関係が認められるもの</p> 	



これらの関係を分類するために, 通常は「しきい値」const2 を定め順序性係数が const2 よりも大きいかな否かによって次の表 4 におけるア, イ, ウの場合に順序性があるとみなすが, ここでは試案として const1 も設定して表 3 のすべての分類に対応させることを考える.

表 4

項目 k から 項目 j から	$\gamma^*_{kj} < \text{const1}$	$\text{const1} \leq \gamma^*_{kj} < \text{const2}$	$\text{const2} \leq \gamma^*_{kj}$
$\gamma^*_{jk} < \text{const1}$	オ (独立) 無印で示す	エ	イ (記号←で示す)
$\text{const1} \leq \gamma^*_{jk} < \text{const2}$	(共通の要因も各々独自の要因も存在) 記号...で示す		
$\text{const2} \leq \gamma^*_{jk}$	ア (記号→で示す)	ウ (記号↔で示す)	

表 4 においてはエが一般の場合を表し, 他は各々顕著な傾向を取り出したものとなっている.

ア, イ, エの場合には, 受け手の項目に送り手になかった独自の要因があることになり, 実際の観測項目の中に該当項目がなければ潜在要因があることになるが, この研究では潜在要因の推定には踏み込まず, 観測項目間の関係のみを扱う.

以下においては経験則から const1 を 0.3 とし, const2 を 0.5 とした.

#### 4. 順序性係数が満たすべき性質の確認

以下の 3 項目のうち, (1)は常に成り立ち(3)は成り立つとは限らないので, 実際上は(2)の「単調性」が実データに対する適合性を判断する上で重要な目安になると考えられる.

(1)  $j \rightarrow k$  のとき相関係数  $r$  や回帰係数  $\beta$  は正になる.

$$\gamma^*_{jk} \geq \text{const2} \text{ のとき, } \gamma^*_{jk} = \beta \frac{p}{q} \geq \text{const2} \text{ となるから } \beta > 0 \text{ となる.}$$

$$\text{また, } \gamma^*_{jk} \geq \text{const2} \text{ のとき, } \gamma^*_{jk} = \sqrt{\frac{p(1-q)}{q(1-p)}} r \geq \text{const2} \text{ となるから } r > 0 \text{ となる.}$$

(2) 一方向の順序関係 ( $j \rightarrow k$ ) のみが認められるときは,  $p > q$  が成立する (単調性). 双方向の関係 ( $j \leftrightarrow k$ ) が認められるときは  $p \doteq q$  となる.

順序関係をその本来の定義「j ができないならば k がで

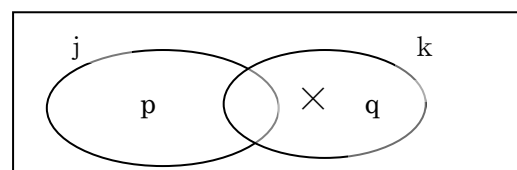


図 2

きない」と解するならば、極端な場合においては図2のように×の部分为空集合となり、 $p > q$ となるはずであるというのが単調性の意味するところであると考えられる。

数式変形としては次のように示せる（筆者の理解）。

$$\gamma_{jk}^* = \frac{ad-bc}{(a+c)(c+d)} \geq \text{const}2 \quad \text{かつ} \quad \gamma_{kj}^* = \frac{ad-bc}{(a+b)(b+d)} < \text{const}2$$

ならば  $\frac{ad-bc}{(a+b)(b+d)} < \text{const}2 \leq \frac{ad-bc}{(a+c)(c+d)}$  だから

$$(a+b)(b+d) > (a+c)(c+d)$$

これを因数分解すると  $b > c$  が示されるから、 $p = \frac{a+b}{N} > q = \frac{a+c}{N}$

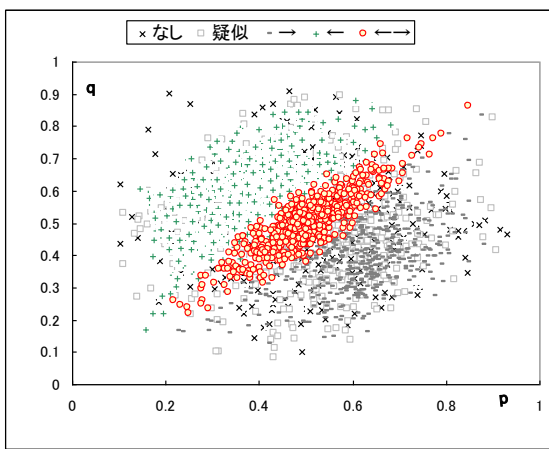


図3

双方向の順序関係が成り立つ場合において、正確な値として  $p = q$  が成り立つ訳ではない

図3はコンピュータ・シミュレーションによって発生させた  $a, b, c, d$  のデータから求められる  $p, q$  の値について、双方向順序関係、 $j \rightarrow k$  の順序関係、 $k \rightarrow j$  の順序関係、疑似相関、順序関係なしの区分ごとに500組示したもので、次の傾向がある。

- $j \rightarrow k$  の順序関係のあるデータについては、 $p > q$  が成り立つ。
- $k \rightarrow j$  の順序関係のあるデータについては、 $q > p$  が成り立つ。
- 双方向の順序関係があるデータは  $q = p$  の直線付近にあり、 $p \doteq q$  となっている。

• 疑似相関および順序関係なしのデータは広く全体的に分布している。

以上のように順序関係は原則として正答率の高い項目から低い項目へ向かうが、2項目の正答率が接近していて双方向順序関係が成り立つときは正答率の低い項目から高い項目にも向かう。

この双方向順序関係が認められる範囲はしきい値によって変化する。  $\text{const}2=0.4$  としたときは図4のように範囲が広くなり、  $\text{const}2=0.7$  としたときは図5のように狭い範囲に集まる。

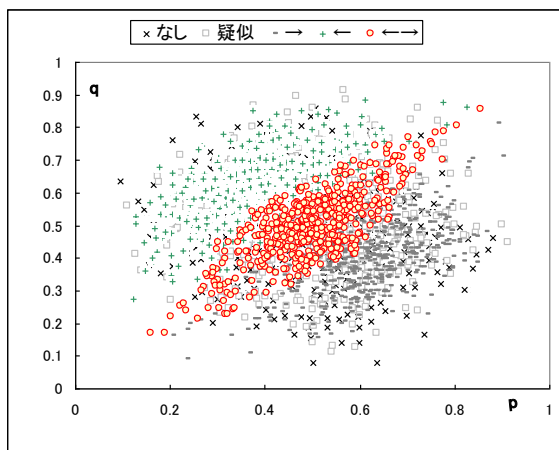


図4

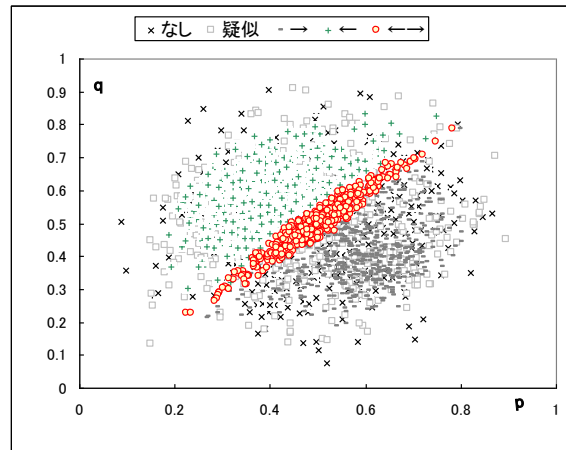


図5

(3) 推移律は必ずしも成り立つとは限らない。  $j \rightarrow k, k \rightarrow l$  のとき2項目間の直接の関係として  $j \rightarrow l$  が成立しているとは限らない。

表5のように項目kが項目jに少しだけ独自の要因を追加しており、項目lも項目kに少しだけ独自の要因を追加してj→k, k→lが成り立つとき、kから見ればjとの共通要因は少なく、必ずしもj→lが成り立つとは限らない。(表5における項目lのように項目jに対して多くの要因を追加すると、孫のlから見ればその要因が親のjから送られたものとは認識できないことがある。したがって、間接的な影響を組み合わせるとつながるはずの関係でも、直接的な数値としては現れないことがある。そこで間接的につながっていればつながっていると考える。)

表5

	項目 j	項目 k	項目 l
要因	x	x	x
要因		y	y
要因			z

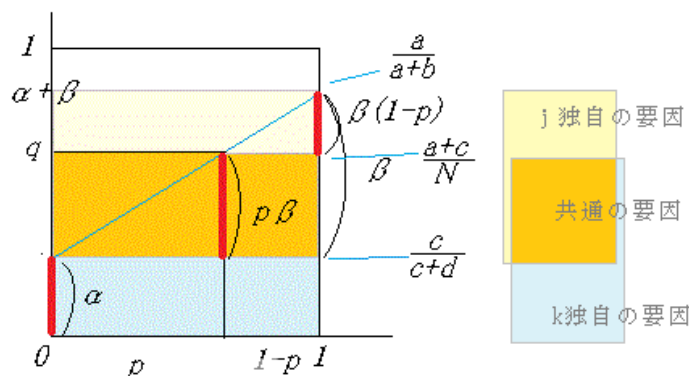


図6

順序性係数  $\gamma^*_{jk} = 1 - \frac{\alpha}{q} = \frac{\beta p}{q}$  がしきい値 const2 以上のとき、「項目jができないならば項目kができない」と解釈されるのが普通であるが、上の図6のように読むと、「項目kを解くために必要な要因のうち共通要因が占める割合」を表しているものと解釈することができる。表5のように様々な要因を追加していくと共通要因の割合は薄くなり順序関係が分からなくなると考えられる。

## 5. 各データが1-0データであるときの実データでの検証

筆者が集計した Web 教材は中学校1年生の数学の問題で、次のような問題計10問から成り立っている。データ件数はN=700である。

第1問	$x+2=5$	第6問	$7x=5x+6$
第2問	$x-4=7$	第7問	$-7-3x=8+2x$
第3問	$3x=6$	第8問	$5(x-3)=4(7-2x)$
第4問	$2x+4=10$	第9問	$-4(x+3)=2(5-x)$
第5問	$4x-8=16$	第10問	$7x-3(x+2)=11$

これらの問題を解くために必要だと考えられる要因を左欄に示し、該当する箇所を1で示すと次の表のようになる。

表6

能力・要因	第1問	第2問	第3問	第4問	第5問	第6問	第7問	第8問	第9問	第10問
定数項移項	1	1		1	1		1	1	1	1
x移項						1	1	1	1	
係数で割り算			1	1	1	1	1	1	1	1
かっこをはずす								1	1	1
解が分数								1		1

この問題を集計したときの相関係数は次の表のとおりであった。

表 7

	正答率	0.829	0.793	0.854	0.803	0.651	0.673	0.456	0.327	0.431	0.324
正答率	相関係数	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
0.829	p1		0.188	0.253	0.270	0.153	0.208	0.165	0.196	..	0.153
0.793	p2	0.188		0.308	0.296	0.284	0.372	0.178	0.176	..	0.151
0.854	p3	0.253	0.308		0.406	0.301	0.333	0.264	0.193	..	0.234
0.803	p4	0.270	0.296	0.406		0.368	0.450	0.288	0.254	0.076	0.274
0.651	p5	0.153	0.284	0.301	0.368		0.423	0.320	0.235	0.117	0.238
0.673	p6	0.208	0.372	0.333	0.450	0.423		0.400	0.350	0.116	0.346
0.456	p7	0.165	0.178	0.264	0.288	0.320	0.400		0.389	0.193	0.396
0.327	p8	0.196	0.176	0.193	0.254	0.235	0.350	0.389		0.241	0.532
0.431	p9	..	..	..	0.076	0.117	0.116	0.193	0.241		0.308
0.324	p10	0.153	0.151	0.234	0.274	0.238	0.346	0.396	0.532	0.308	

これによりプログラムから出力される順序関係表は次の表のとおり。

表 8

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
p1						...	...	→		...
p2			...	...	...	→	...	...		...
p3				...	→	→	→	→		→
p4					→	→	→	→		→
p5						...	...	...		...
p6							→	→		→
p7								→		→
p8									...	←→
p9										...
p10										

表 8 により順序関係→, ←, ←→が成り立つものを示し, すでに間接的に結ばれている項目間には新たな矢印は書かない. 擬似相関または弱い順序関係であるときに潜在要因を推定すると複雑になりすぎるため, すでに順序関係で結ばれている観測項目間の関係を目で確認する程度とする. このとき次の図のようなグラフが得られる. なお, 第 9 問はよく考えなければできない選択問題において先頭の選択肢を正答としたため, まぐれあたりが多く発生し, 関係が切れてしまったものと考えられる.

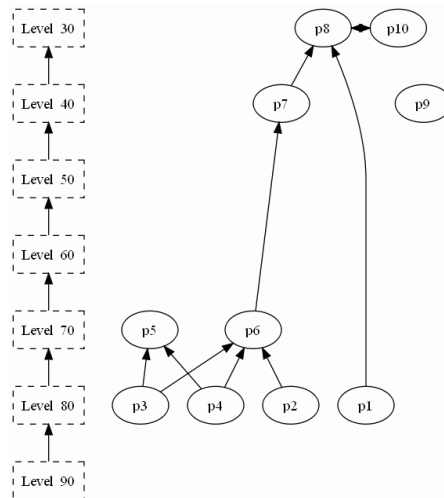


図 7

なお参考までに、誤答原因を調べるために他の分析方法を試みた結果は、概略次の通りであった。

(1) 主成分分析

第5主成分まで求めて係数の絶対値の大きな項目を1で示したときは次の表のようになり、能力・要因との関連はよく分からない。

表9

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	備考
主成分1			1	1	1	1	1	1		1	係数で割ること？
主成分2		1						1	1	1	該当要因不明
主成分3	1				1						該当要因不明
主成分4								1	1		該当要因不明
主成分5		1	1				1				該当要因不明

(2) 因子分析

4因子・回転なしで因子分析を行ったときの独自因子、因子負荷量は次の表のようになり、第2問 (p2) 以外は独自要因が大きくなり過ぎて、能力・要因などの関連はよく分からない。

表10

独自因子	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
	0.85	0.005	0.665	0.477	0.645	0.489	0.608	0.518	0.846	0.409
因子負荷量	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
因子1	0.269		0.402	0.504	0.425	0.548	0.575	0.605	0.337	0.658
因子2	0.159	0.997	0.29	0.297	0.283	0.379	0.182	0.19		0.164
因子3	0.115		0.268	0.384	0.223	0.191		-0.284	-0.196	-0.347
因子4	0.183		0.135	0.182	-0.211	-0.172	-0.17			0.1

(3) 重回帰分析

表9のようにあらかじめ誤答要因が想定できる場合に、各誤答要因の影響の大きさを確認的に分析するとよい結果が得られることがあるが、項目間の関連を探索的に分析するのは難しい。

## 6. 各データが正答率であるときの順序性係数

6.1 1.で示した1-0データは次のようなものであった。

表1 (再掲)

項目 j	1	1	1	1	1	1	1	1	0	0	0	0	0	0
項目 k	1	1	1	1	1	0	0	0	1	1	1	1	0	0
集計	a=5					b=3			c=4			d=2		

各データが正答率であるとき、次の表11のようなデータになる。

表11

項目 j	0.9	0.8	0.7	0.6	0.8	0.8	0.6	0.7	0.1	0.2	0.3	0.1	0.0	0.1
項目 k	0.7	0.8	0.6	0.9	0.7	0.1	0.2	0.3	0.6	0.7	0.7	0.8	0.1	0.2

もともと、順序性係数  $\gamma^*_{jk}$  は1-0データを集計した  $2 \times 2$  分割表の

観測度数 a, b, c, d を用いて

$$\gamma^*_{jk} = 1 - \frac{cN}{(c+d)(a+c)} \quad (\text{竹谷 1979}) \text{ と定義されており, そ}$$

もそも正答率で与えられるデータに対して順序性係数が定義できるのかという理論的な問題がある。この問題に対する先行研究が見つけれなかったので、ここでは2つの試案とこれを実データに適用した場合の結果を示す。

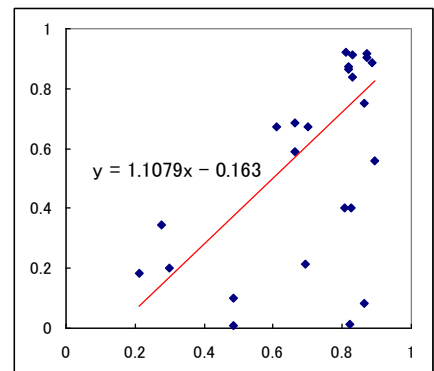


図8



(1) 正答率の小数值のまま順序性係数を表す式を考える場合

データが正答率である場合には、回帰係数  $\beta$  は次の関係を満たす。

$$\beta = \frac{s_k}{s_j} r = \frac{s_k s_{jk}}{s_j s_j s_k} = \frac{s_{jk}}{s_j^2} \dots(17)$$

これにより

$$\gamma_{jk}^* = \beta \frac{p}{q} = \frac{p s_{jk}}{q s_j^2} \dots(18)$$

を順序性係数の定義式と考える。ただし、順序性係数の有意性の判断は相関係数の有意性によって行う。

$$\gamma_{jk}^* = \beta \frac{p}{q} = \frac{p s_k}{q s_j} r \dots(19)$$

この場合には、右図のように回帰直線の定数項  $\alpha$  が負の値になる場合や回帰係数  $\beta$  が 1 よりも大きな値になる場合も起こるので、改めて順序性係数が表している内容を考えなければならなくなるが、その問題は保留として実データへ適用したときの適合性を以下において調べた。

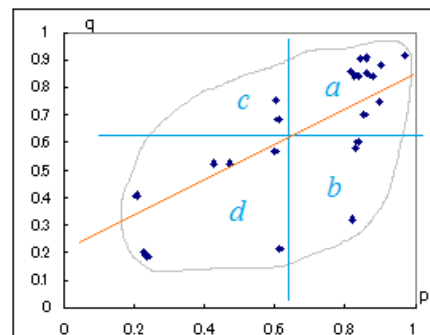


図 9

(2) 正答率で与えられるデータを四捨五入などの方法により 1-0 データにデジタル化する場合

表 11 のような正答率データは四捨五入などの方法によって 1-0 データに直すことができる。この他、各頁の平均点よりも上か下かで分ける方法、目標準拠的な取り扱いとして設定通過率を 70% などとしてこれよりも上か下かという考え方もありうる。いずれも何らかのボーダーラインを設定して 2×2 分割表に当てはめれば順序性係数の本来の定義式を適用できることになる。

6.2 順序関係の検出率の比較

次のグラフは、20 項目からなる実データの  ${}_{20}C_{19} \times 2 = 380$  通りの組み合わせについて「小数值のまま順序性係数を計算した場合」、「ボーダーラインを 0.5, 0.55, ..., 0.85 に設定した場合」、「問題ごとにその平均点をボーダーラインに設定した場合」について順序関係の検出率を調べたものである。(近似曲線は 2 次曲線とした。) 各項目に一つの矢印が入っており、かつ 1 つの矢印が出ている場合には検出率  $40/380 = 0.10$  となるので各項目間に何らかの順序関係を読み取ることができると考えられ、検出率が 0.1 以上あれば関連付けを行うことができる。

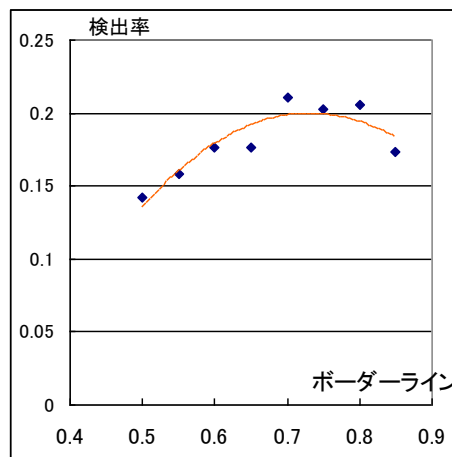


図 10

図 10 のグラフによれば全体の平均正答率 66% よりも少し大きな値 70% にボーダーを設定したときに、順序関係が最も検出されやすいことが分かる。

なお、各頁の平均点以上か否か (頁ごとに異なる基準) で分けた場合は (グラフ中の点として示されていない)、その検出率は 0.13 と非常に低くなる。また、小数值のまま順序性係数を求めた場合には 0.17 になった。

同様に、次のグラフは中学校 2 年～3 年生の 19 項目について、ボーダーラインと順序関係検出率を調べたもので、この場合にも全体の平均正答率 71% よりも少し大きな値 75%～80% で分けたときに検出率が最大となる。なお各頁の平均点以上か否か (頁ごとに異なる基準) によって項目間の関連を検出したときの検出率は 0.10 と低い値になり、また、小数值のまま順序性係数を求めた場合には 0.05 と低い値

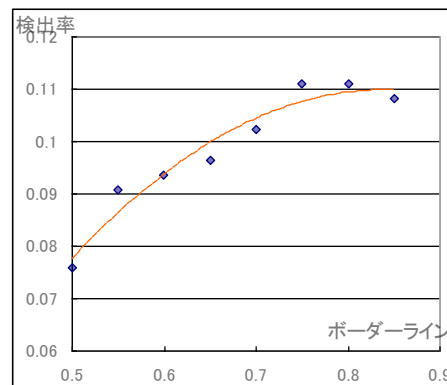


図 11

になった。

### 6.3 順序関係の検出率が変わる原因

右図 12 において項目 k が項目 j と無関係な場合には  $q = \alpha$ ,  $\beta = 0$  となって楕円の上半分と下半分は同じ面積になる。これに対して傾き  $\beta$  を大きくすると  $\alpha/q$  が小さくなり、楕円の左半分のうち上半分の割合が減少する。これが「j ができない者は k もできない」ということであると考えられる。

ここで扱ったデータでは、右図 12 のような形に分布することが多いので各々の平均値よりもやや大きな値をボーダーラインに選ぶと左上に入る点が少なくなり、順序関係が「過剰に」検出されると考えられる。

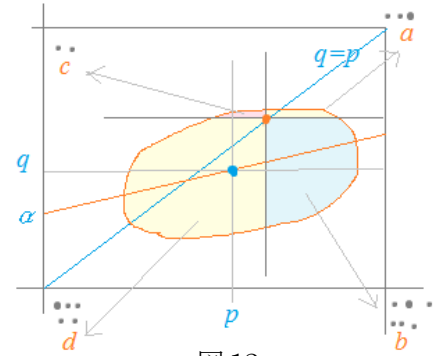


図 12

## 7. 各頁の得点が正答率であるときの実データでの検証

6. で述べた方法によって、各頁の得点が正答率で与えられたときに順序関係を求めた例を示す。使用したデータは、2010.2.1 から 2010.6.15 までの期間に筆者の Web 教材に回答のあった 22,282 件の答案のうち同一人物の答案を除いて 9,410 人の回答を選び、その中から対応する 2 項目の両方に回答している答案を抽出したものである。（\*このデータについての補足説明は末尾の参考資料）

表 12 は中学校 1 年生の単元の一部を示したもので、表中の記号は表 4 の約束により、独立であるものは無印、左欄の項目から上欄の項目への順序関係が見られるものは  $\rightarrow$ 、上欄から左欄への順序関係が見られるものは  $\leftarrow$ 、双方向順序関係が見られるものは  $\leftrightarrow$ 、擬似相関（共通の原因を持つものまたは弱い順序関係）は  $\dots$  で示した。

(1) 1・0 データに変換するためのボーダーラインを 0.7 に設定した場合

表 12

	p18	p21	p22	p24	p28	p31	p34	p38	p41
p18		$\leftarrow$			$\dots$			$\dots$	$\dots$
p21				$\dots$	$\leftrightarrow$	$\dots$	$\dots$	$\dots$	$\leftrightarrow$
p22				$\rightarrow$	$\dots$	$\rightarrow$	$\dots$	$\rightarrow$	$\dots$
p24						$\dots$	$\dots$	$\rightarrow$	$\dots$
p28						$\dots$	$\rightarrow$	$\rightarrow$	$\rightarrow$
p31							$\dots$	$\rightarrow$	$\leftarrow$
p34								$\rightarrow$	$\dots$
p38									$\leftarrow$
p41									

表 12 に登場する頁が取り扱っている単元は表 13 の通りで、観測項目の順序関係を図示し疑似相関は確認のみに使用すると次の図のようになる。矢印が複雑になるのを避けるため、すでに間接的に結ばれているときは新たな矢印は書かない。

表 13

p18	文字の使用
p21	文字を使った式 [特訓]
p22	文字を使った式
p24	文字と式・速さ・時間・距離
p28	式の値 1
p31	関係を表わす式
p34	方程式
p38	方程式 (分数係数)
p41	中一・1 次方程式 (ゲーム 1)

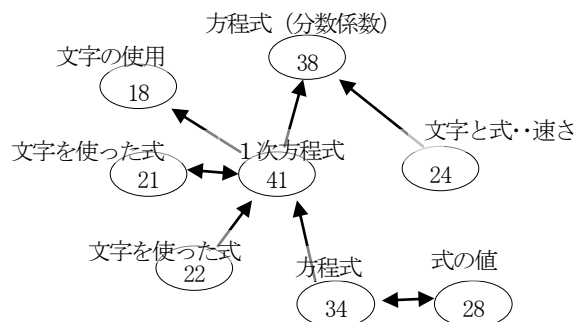


図 13

同様にして、中学校1年生のx y座標, 比例・反比例, 空間図形については次のグラフが得られる。

表 14

p50	x y座標→点1
p51	x y座標→点2
p53	比例1
p54	比例2
p55	反比例のグラフ1
p57	反比例
p60	比例・反比例のグラフ
p64	立体の体積
p67	扇形の面積
p68	立体の表面積

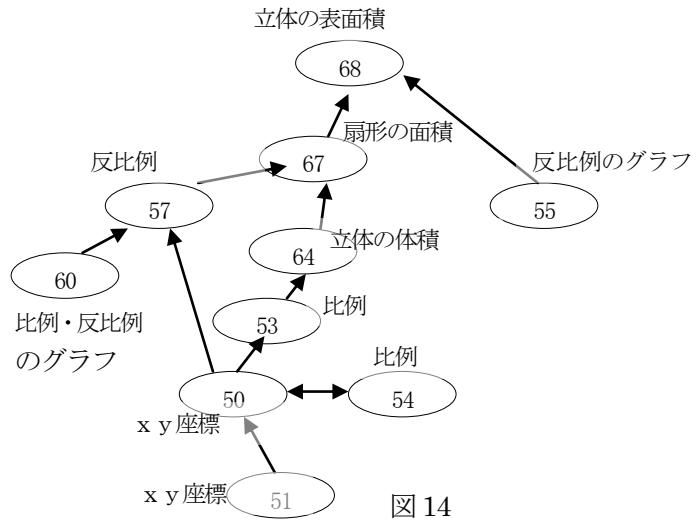


図 14

(2) 正答率の小数值のまま順序性係数を表す式を考える場合

図 13, 図 14 と同じ項目は次のような関連構造をもつと解釈される。

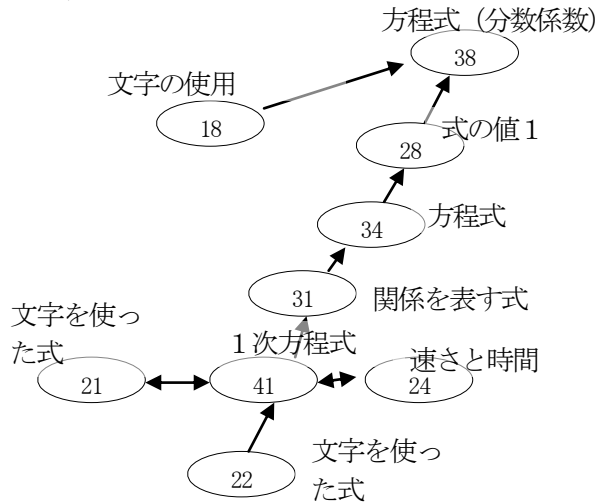


図 15

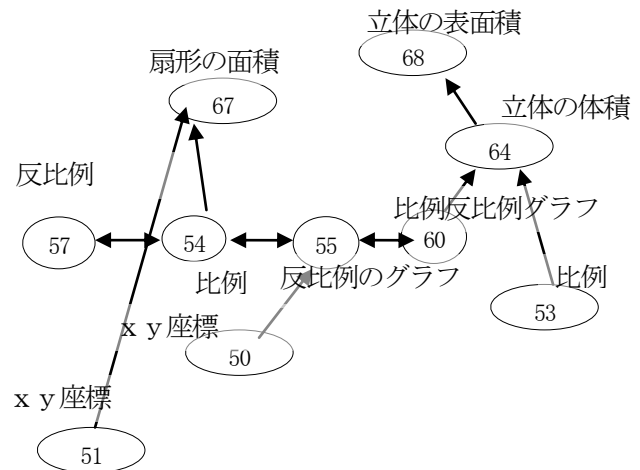


図 16

図 13 と図 15, 図 14 と図 16 を比較してみると項目間の関連について異なる解釈が行われていることになる。印象だけで述べれば「正答率の小数值のまま順序性係数を求める」方が優れているようであるが、何らかの客観的な基準によって実データへのモデルの適合性を判断する方がよいと考えられる。

(以下に、幾つか具体例を示す。)

・図 13 において、「1次方程式」から「文字の使用」へ順序関係があるのは常識に反する。これに対して図 15 においては、概ね「文字式」→「方程式」の順序関係となっており常識と一致する。

・図 14 において、「立体の体積」から「扇形の面積」へ順序関係が認められるのは常識に反する。

この答案の散布図を示すと右図のようになり(整数値の重なりを避けるため揺らぎを加えている)、「立体の体積」の項目の正答率は  $p=0.529$  , 「扇形の面積」の項目の正答率は  $q=0.356$  , 相関係数は  $0.355$  , 回答者数  $N=60$  である。

右図 17 において「立体の体積」ができない者(左半分)のうち「扇形の面積」ができた者(上半分)が特に少ないとはいえず、小数值のまま順序性係数を求めると  $0.31$  となり、順序性は認められ

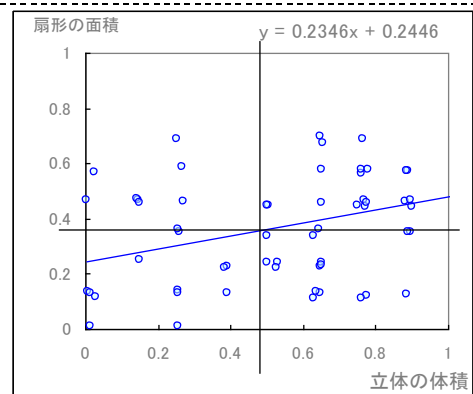


図 17

ない。

回帰直線で見るときはqに対してαの位置が高過ぎるので、順序性が認められないと考えられる。

これに対して、ボーダーラインを0.7として1-0データに変換すると、左半分のうち上半分に入るデータは0個となり順序性係数が1、すなわち強い順序関係があると解釈していることになる。

右の図18は、「立体の体積」→「立体の表面積」の答案の散布図(p=0.566, q=0.361, r=0.359, N=85)で、小数値のまま計算して順序関係が認められるものとなっている。(整数値の重なりを避けるため揺らぎを加えている)

このグラフでは左半分のうち上半分は「薄く」なっており、「立体の体積」ができないならば「立体の表面積」ができないと言える。

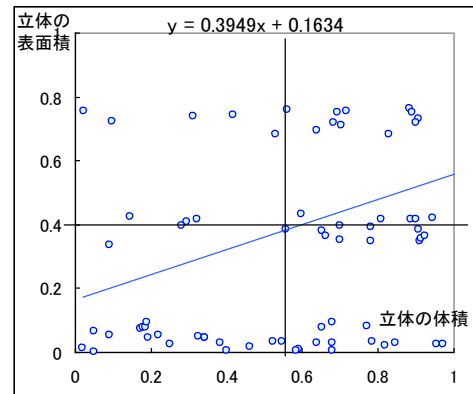


図18

回帰直線で見るときはqに対するαの位置が十分低いことから分かる。

なお4.で述べた「単調性」について、j→kの順序性が求められるとき、ほとんどの場合p>qが成り立つが、まれにp/s<sub>j</sub>>q/s<sub>k</sub>のみ成り立ちp>qとならないことがある。(p=0.6469, q=0.7188, r=0.4579, s<sub>j</sub>=0.2610, s<sub>k</sub>=0.3273 → γ<sub>jk</sub>\*>const2)

\*参考資料1 [単調性について]

7節以降において「各頁の得点が正答率であるときの実データ」として用いたものは、Web教材に対する断片的な回答をつないだものであり、一般のテスト結果と異なり「すべての回答者がすべての問題を解いている」訳ではない。

右に例を示すように、1頁と2頁の関係を調べているときには、これら両方の頁に回答している者から各々の正答率p1, p2を求めており、2頁と3頁の関係を求めているときにはこれら両方の頁に回答している者から正答率p2', p3'を求めているので、p2=p2'とは限らない。

1頁	2頁	3頁
60	50	
40		60
...	...	...
	75	65

このため「1頁」→「2頁」、「2頁」→「3頁」の順序関係が認められた場合において、各正答率間の関係としてp1>p2, p2'>p3'のように2項目間の単調性が成り立つことを確かめることはできるが、「1頁」と「3頁」の関係においてはp1<p3'となることがある。

回答者数が多くなれば問題ごとの正答率は安定してくるが、正確には一致しない。

\*参考資料2 [グラフ作成ソフトについて]

Excel上で、例えば右の表15のようにテキストとして出力されるようにしておき、これをgraphvizというフリーソフト(<http://www.graphviz.org/>)の編集画面で次のように貼り付けると、項目関連グラフが自動的に描画され、gifファイルとして出力される。

表15

```
digraph ex1
{
    [この箇所]
}
```

```
p1 -> p3 [dir=back];
p2 -> p3 [dir=both];
p2 -> p4;
p3 -> p4;
p3 -> p6;
p5 -> p6 [dir=both];
p4 -> p8 [dir=back];
p5 -> p8 [dir=back];
p6 -> p8 [dir=back];
p3 -> p9;
p4 -> p9;
p8 -> p9;
```

ただし、このグラフ作成ソフトでは、項目間の矢印は表15に忠実に結ばれ、表15において「間接的な順序関係」と「直接的な順序関係」に重複があるときに「直接的な順序関係」が自動的に取り除かれる訳ではない。したがって、図19のグラフを「目で見て」不要な矢印、例

例えば `p3 -> p9;` などを表15から取り除いてから再度走らせると図20が得られる。

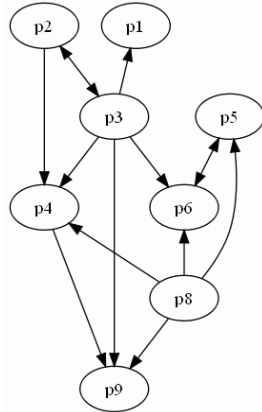


図 19

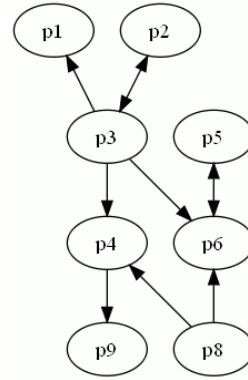


図 20

また、このグラフ作成ソフトは初期設定において項目関連グラフの約束とは逆に上から下へ矢印が引かれる。そこで、例えば正答率 50%, 60%, 70%のレベルを Level\_50, Level\_60, Level\_70 などで表示し、各項目をレベルに応じてランク付けすると通常の項目関連グラフと同様に、正答率の高い項目は下に、低い項目は上に、正答率の近い項目は横に並ぶようにできる。(図 19, 20 では表示できなかった独立な項目 p7 もグラフの中に入れてられる。)

```

digraph ex1
{
level_50 -> level_60 -> level_70 [dir=back];
level_50[shape=box];
level_60[shape=box];
level_70[shape=box];
p5 -> p8 [dir=back];
p5 -> p6 [dir=both];
p4 -> p9;
p4 -> p8 [dir=back];
p3 -> p6;
p3 -> p4;
p2 -> p3 [dir=both];
p1 -> p3 [dir=back];

{rank=same; level_50;p1;p6;p9;p7}
{rank=same; level_60;p4;p5}
{rank=same; level_70;p2;p3;p8}
}

```

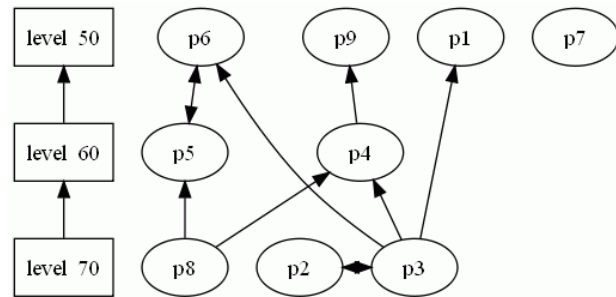


図 21

p78	多項式の和差
p98	1次関数(傾き, 切片)
p106	外角
p107	図形1(対頂角, 同位角, 錯角)
p111	証明の進め方
p112	証明の進め方3
p113	円周角の定理
p115	円に内接する四角形
p116	円周角の定理・類題
p120	確率の求め方
p124	乗法の公式2
p125	乗法の公式1
p135	因数分解(共通因数→)
p137	因数分解(→2乗の形)
p138	素因数分解
p140	素数
p142	2乗・平方根(用語)
p145	2次方程式の解き方
p149	三平方の定理

左のような2, 3年生の単元について graphviz を用いて項目関連グラフを作成すると次のようになる。ただし、グラフはできるが範囲が広過ぎるため内容的な関連は必ずしも納得できるものではない。

