

人の感覚に合った最適距離のパラメータ推定

小笠原 悠 (Yu OGASAWARA)

金 正道 (Masamichi KON)

弘前大学大学院 理工学研究科

Graduate School of Science and Technology, Hirosaki University

概要

本研究は、人間の感覚に合った距離を良い距離とし、いくつかのパラメータで定められる距離の中で最適な距離を実現するパラメータを、数量化 III 類を行った分析結果を事例を用いて算出した。また、その算出されたパラメータを確率変数と捉え、パラメータ分布の統計量をブートストラップ法を用いて推定し、各パラメータの特徴を観測した。

1 はじめに

構造記述モデル分析法として主成分分析や数量化 III 類などが知られている。その分析結果に対して 2 点間の距離を定義し、クラスター分析や最適施設配置問題などを用いてさらに分析を進めることがある。そのようなときに、距離をどのように定義するかによって分析結果が異なる場合があり距離の選択が重要になる。問題の種類によってさまざまな距離が考えられているが、意思決定者の意図に合った解釈ができる距離を選択したり、分析の進め易さ（計算の容易さ）や分析結果の解釈のし易さによって距離選択がなされる場合が多い。

Love and Dowling[4] や Love and Morris[5] は実際の道路に対して、最適な距離関数の近似を行っている。しかし、道路に対しての最適な距離選択は行われているが、人間の感覚的距離に対して適用されている研究は無い。そこで本研究は人間の感覚に対して最適な距離選択を分析することを試みた。本研究では人間の感覚に合ったものをよい距離と考える。ある 2 つの対象（人やもの）があった場合、人の判断によってその 2 つの対象間の距離を定めることは困難であったり、判断する人によってばらつきが大きいことが予想されるが、ある 2 つの対象間の距離は他の 2 つの対象間の距離に比べてどのぐらいかという比を判断することは比較的容易であり、距離を直接定める場合よりは判断する人によってのばらつきも小さいと予想する。ここでは、後者の考え方に従って、対象間の距離を一対比較することによって、一対比較行列を作成し、AHP を適用して得られたウェイトを用

いて対象間の距離の比を構成し、その比と距離関数によって測られた対象間の距離の比を比較して、誤差が小さくなるような距離（関数）を人間の感覚に合ったものとし、よい距離とする。対象間の距離を一対比較ではなく、直接対象間の距離を判断した場合、判断ミスによって、人間の感覚に合った距離と異なる誤った選択をしてしまう可能性が大きいと予想されるが、対象間の距離を一対比較してAHP適用することによってある程度の判断ミスの発見や修正が期待できる。

本研究は、人間の感覚に合った距離を良い距離とし、いくつかのパラメータで定められる距離の中で最適な距離を実現するパラメータを、数量化III類を行った分析結果を事例を用いて算出した。また、その算出されたパラメータを確率変数と捉え、パラメータ分布の統計量をブートストラップ法を用いて推定する。第2章で事例とする分析結果と、本研究で使用する距離関数とパラメータを紹介する。第3章で本研究で行ったアンケートと計算結果を示す。第4章で算出された最適な距離を実現するパラメータの分布をブートストラップ法を用いて推定し、第5章にて結論を述べる。

2 最適距離選択

2.1 青森県による事例とカテゴリストア

生涯学習・社会教育に関する調査研究として、青森県在住の成人を対象にアンケートによる家庭教育に関する意識調査が青森県総合社会教育センターによって行われた。この調査は、家庭の教育力を充実するために、県民が家庭教育に関する学習内容や学習活動等に対して、どのような要求課題を持っているかを明らかにし、市町村教育委員会などの各学習提供機関の基礎資料として提出する目的で行われ、調査結果が報告された [1]。M. Kon[3]はこのアンケート結果に数量化III類を適用して分析を行っている。ここでは、下に載せる質問項目に対するアンケート結果を用いて、「乳幼児の時期の子どもにとって必要な教育項目」を考える。

アンケートの質問項目

問 家庭において、お子さんが乳幼児・小学生・中学生・高校生の時期に、もっとも重要だと思われる教育項目は何ですか。次の中から最大3つまで選び番号を記入してください。（同じ項目を何回選んでもかまいません）

- 1 基本的な生活習慣（例えば、洗顔、自分で起床、あいさつなど）
- 2 生活体験（例えば、タオルをしぼる、小さな子の世話、ナイフの使い方など）
- 3 自然体験（例えば、海や川で遊ぶ、自然観察、登山など）
- 4 自主性（自分の判断で行動する態度）
- 5 自制心（感情・欲望などを自分で抑えること）
- 6 自立心（人に頼らず、独り立ちして自力でやっという心構え）

- 7 豊かな情操（美しいものを美しいと感じる心）
- 8 他人への思いやり
- 9 道德感
- 10 社会的なマナー
- 11 正義感
- 12 人間関係づくり
- 13 職業観
- 14 性教育
- 15 その他（具体的にお書きください）
- 16 わからない

得られたアンケート結果から，反応数が少なかったカテゴリ（回答項目）を除き

- 1 基本的生活習慣 2 生活体験 3 自然体験 7 豊かな情操

のみを分析対象とした．欠損がある個体を除いて表1のようなカテゴリ・データが得られた．表2は，表1のアンケート結果に数量化 III 類を適用した結果の固有値を示している．ここでは，第2位までの固有値を取り上げる．表3は，カテゴリスコアを示している．カテゴリスコアは次のように解釈できる．カテゴリ第1スコア z_1 の値が小さいほど身心的（情操的）な事に関する学習内容を表し，大きいほど身体的（行動的）な事に関する学習内容を表す．カテゴリ第2スコア z_2 の値が小さいほど（場所が）非日常的な事に関する学習内容を表し，大きいほど（場所が）日常的な事に関する学習内容を表す．カテゴリ番号1, 2, 3, 7の各カテゴリスコアをそれぞれ y_1, \dots, y_4 とする．

表1: アンケート結果

個体番号	1 生活習慣	2 生活体験	3 自然体験	7 豊かな情操
1	1	0	0	1
2	1	1	1	0
⋮	⋮	⋮	⋮	⋮
492	1	1	1	0

表2: 固有値

順位	固有値	寄与率	累積寄与率	相関係数
1	0.357	53.28	53.28	0.598
2	0.176	26.29	79.58	0.420
3	0.137	20.42	100.00	0.370

表3: カテゴリスコア

カテゴリ番号	(z_1 : 第1スコア, z_2 : 第2スコア)
1 (生活習慣)	(0.071, 0.315) = y_1
2 (生活体験)	(1.368, 1.128) = y_2
3 (自然体験)	(0.343, -1.721) = y_3
7 (豊かな情操)	(-1.790, 0.414) = y_4

2.2 距離関数とその他指標について

次に, y_i と y_j の間の感覚的な距離 ($i-j$ と表す) を一対比較するアンケートを行い, 例として表4のような一対比較行列が得られたとする. 表4の一対比較行列に対して AHP を適用すると表5のようなウェイトが得られる. この1-2, 1-3, ..., 3-4のウェイトを順に w_1, \dots, w_6 とする. 得られたウェイトは比にのみ意味があるため, それらの比を感覚的な距離の真の比 (理想的な比) と見なし, 距離関数の選択基準に利用する. 表5から感覚的な距離の真の比を算出したものが表6になる. 表6によって与えられる行列の i 行 j 列の値は $\frac{w_i}{w_j}$ の値である.

表4: 一対比較行列

	1-2	1-3	1-4	2-3	2-4	3-4
1-2	1	$\frac{1}{7}$	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{9}$
1-3	7	1	1	3	5	1
1-4	3	1	1	1	1	$\frac{1}{5}$
2-3	5	$\frac{1}{3}$	1	1	5	$\frac{1}{3}$
2-4	3	$\frac{1}{5}$	1	$\frac{1}{5}$	1	$\frac{1}{7}$
3-4	9	1	5	3	7	1

表5: AHP によって得られたウェイト

	ウェイト
1-2	0.031 = w_1
1-3	0.269 = w_2
1-4	0.118 = w_3
2-3	0.147 = w_4
2-4	0.064 = w_5
3-4	0.371 = w_6

表6: 感覚的な距離の真の比 (理想的な比)

	1-2	1-3	1-4	2-3	2-4	3-4
1-2	1.000	0.116	0.262	0.211	0.488	0.084
1-3	8.644	1.000	2.268	1.822	4.219	0.724
1-4	3.811	0.441	1.000	0.803	1.860	0.319
2-3	4.743	0.549	1.245	1.000	2.315	0.397
2-4	2.049	0.237	0.538	0.432	1.000	0.172
3-4	11.946	1.382	3.134	2.518	5.830	1.000

ここで、 $\mathbf{y}_i = (y_i^1, y_i^2)$ と $\mathbf{y}_j = (y_j^1, y_j^2)$ の間の距離を測る際に、本研究では、人は各カテゴリストアの要素を個人感覚で歪ませた直角距離によって感覚的な距離を求めていると仮定する。つまり、距離関数は以下になる。

$$d_{a,b,\theta}(\mathbf{y}_i, \mathbf{y}_j) = a|y_i^1(\theta) - y_j^1(\theta)| + b|y_i^2(\theta) - y_j^2(\theta)| \quad (1)$$

ここで、 $a, b > 0, 0 \leq \theta < \frac{\pi}{2}$ であり

$$\begin{pmatrix} y_i^1(\theta) & y_i^2(\theta) \\ y_j^1(\theta) & y_j^2(\theta) \end{pmatrix} = \begin{pmatrix} y_i^1 & y_i^2 \\ y_j^1 & y_j^2 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (2)$$

である。 $a, b > 0$ および $0 \leq \theta < \frac{\pi}{2}$ に対して、 $d_{a,b,\theta}$ は $d_{1,1,0}$ の単位円を x 軸の正方向に対して半時計回りに θ だけ回転させ、 x 軸方向に $\frac{1}{a}$ 倍、 y 軸方向に $\frac{1}{b}$ 倍したものを単位円としてもつような距離である。

ここで、 $d_{a,b,\theta}, a, b > 0, 0 \leq \theta < \frac{\pi}{2}$ を任意に 1 つ固定し、 $d = d_{a,b,\theta}$ とする。このとき

$$\begin{aligned} \tilde{w}_1 &= d(\mathbf{y}_1, \mathbf{y}_2) = d(\mathbf{y}_2, \mathbf{y}_1), & \tilde{w}_2 &= d(\mathbf{y}_1, \mathbf{y}_3) = d(\mathbf{y}_3, \mathbf{y}_1), \\ \tilde{w}_3 &= d(\mathbf{y}_1, \mathbf{y}_4) = d(\mathbf{y}_4, \mathbf{y}_1), & \tilde{w}_4 &= d(\mathbf{y}_2, \mathbf{y}_3) = d(\mathbf{y}_3, \mathbf{y}_2), \\ \tilde{w}_5 &= d(\mathbf{y}_2, \mathbf{y}_4) = d(\mathbf{y}_4, \mathbf{y}_2), & \tilde{w}_6 &= d(\mathbf{y}_3, \mathbf{y}_4) = d(\mathbf{y}_4, \mathbf{y}_3) \end{aligned} \quad (4)$$

とし、 d の感覚的な距離からの誤差として次のようなものを考える。

$$E_1(d) = \sum_{i,j} \left(\frac{\tilde{w}_i}{\tilde{w}_j} - \frac{w_i}{w_j} \right)^2 = \sum_{i \neq j} \left(\frac{\tilde{w}_i}{\tilde{w}_j} - \frac{w_i}{w_j} \right)^2 \quad (5)$$

$$E_2(d) = \sum_{i,j} \left(\log \frac{\tilde{w}_i}{\tilde{w}_j} \right)^2 = 2 \sum_{i < j} \left(\log \frac{\tilde{w}_i}{\tilde{w}_j} \right)^2 \quad (6)$$

$$E_3(d) = \sum_{i,j} \left(\frac{\tilde{w}_i}{\tilde{w}_j} - \frac{w_i}{w_j} \right)^2 = \sum_{i \neq j} \left(\frac{\tilde{w}_i}{\tilde{w}_j} - 1 \right)^2 \quad (7)$$

$E_1(d)$ はウェイト毎の誤差の和を取ったものになり、 $E_2(d)$ はそれに対数を取ったもの、 $E_3(d)$ は誤差を率にしたものになっている。以降、 $E_1(d), E_2(d), E_3(d)$ は E_1, E_2, E_3 と表記する。

また、 $a, a', b, b', \lambda > 0$ に対して

$$a' = \lambda a, \quad b' = \lambda b$$

ならば

$$d_{a',b',\theta}(\mathbf{y}_i, \mathbf{y}_j) = \lambda d_{a,b,\theta}(\mathbf{y}_i, \mathbf{y}_j)$$

となるので、誤差 $E_k, k = 1, 2, 3$ の定義より各 $k = 1, 2, 3$ に対して

$$E_k(d_{a',b',\theta}) = E_k(d_{a,b,\theta})$$

となることに注意する。よって、 $a, b > 0$ は

$$(a, b) = (\cos \eta, \sin \eta), \quad 0 < \eta < \frac{\pi}{2} \quad (8)$$

の範囲を考えれば十分であるので、誤差を最小にするようなパラメータ a, b を探索するときは η に関して $0 < \eta < \frac{\pi}{2}$ の範囲を探索すれば十分である。

よって、事例から求められたカテゴリストアと、その質問項目に対する一対比較するアンケート結果から、 $d_{a,b,\theta}$ の距離関数に対して、 E_1, E_2, E_3 の各誤差を最小にする最適なパラメータ θ, η が、アンケートを行った個体毎に得られることになる。次の章で、本研究にて行ったアンケートと、その計算結果を述べる。

3 最適パラメータの計算・分析結果

この章ではパラメータを算出するために本研究で行ったアンケートと、そのアンケートを用いて計算を行った結果を紹介する。本研究では、青森で行われたアンケート事例 [1] における「乳幼児の時期の子どもにとって必要な教育項目」で、特に反応数の多かった、基本的な生活習慣、生活体験、自然体験、豊かな情操の一対比較するアンケートを行った。得られる一対比較行列は先の章の表 4 の形になる。アンケートは 108 人を対象に行ったが、AHP を適応する際に整合度が 1.5 以上になった個体に関しては計算対象から除外した。その結果、計算対象となった個体は 47 人となった。本研究ではこの 47 人分の表 6 のような感覚的な距離の真の比を用いて、最適な距離を実現するパラメータ θ, η を算出した。計算は C++ で組んだプログラムにてステップ幅 0.001 の格子点探索法で行った。 $d_{a,b,\theta}$ に対して、 E_1, E_2, E_3 の各誤差を最小にするパラメータ θ, η の一覧を表 7 に載せる。

表 7: $d_{a,b,\theta}$ に対する最適なパラメータ

個体番号	η	θ	E1	η	θ	E2	η	θ	E3
0	0.654	0.202	17.486	0.687	0.157	13.614	0.691	0.133	25.196
1	1.435	1.118	138.173	1.410	1.117	39.614	1.359	1.184	170.975
2	0.006	0.482	38.901	0.001	0.454	19.389	0.001	0.422	39.026
3	0.238	1.517	173.256	0.442	1.517	39.100	0.485	1.517	138.000
4	1.251	0.211	37.641	1.228	0.184	19.406	1.243	0.140	42.328
5	0.315	0.785	57.096	0.415	0.785	21.387	0.459	0.785	43.125
6	0.001	0.589	66.290	0.001	0.576	15.886	0.001	0.584	33.033
7	0.019	0.509	238.454	0.003	0.503	65.161	0.631	0.222	617.207
8	0.001	0.399	43.343	0.001	0.422	13.504	0.001	0.427	22.707
9	0.001	1.253	96.543	0.001	1.239	26.865	0.001	1.235	86.400
10	0.733	0.785	23.590	0.622	0.785	11.972	0.638	0.785	18.154
11	0.277	0.133	54.811	0.285	0.133	21.575	0.358	0.166	52.109
12	1.570	0.714	209.534	1.570	0.821	27.272	1.547	0.869	72.556
13	1.400	0.222	85.804	1.415	0.222	29.717	1.381	0.222	70.653
14	1.209	1.122	15.170	1.278	1.098	10.573	1.256	1.097	15.407

15	0.251	1.027	17.285	0.181	1.029	12.300	0.125	1.029	21.041
16	0.449	0.785	19.545	0.480	0.785	11.409	0.502	0.785	17.586
17	1.236	0.289	17.831	1.293	0.302	9.336	1.281	0.304	13.501
18	1.260	0.333	15.572	1.245	0.352	6.901	1.241	0.356	8.878
19	1.393	0.810	69.834	1.015	0.785	22.139	1.009	0.785	46.861
20	1.529	0.938	31.003	1.517	0.963	10.718	1.492	0.977	16.427
21	0.451	0.785	51.902	0.459	0.785	30.431	0.626	0.785	87.717
22	0.197	0.875	23.623	0.141	0.880	13.453	0.107	0.901	22.491
23	1.018	1.225	15.927	0.946	1.225	8.951	0.916	1.225	13.328
24	0.078	0.993	28.848	0.125	0.989	13.724	0.113	1.012	22.040
25	0.465	1.008	9.222	0.413	1.002	7.943	0.389	1.026	10.907
26	1.353	0.560	72.621	1.329	0.560	15.148	1.264	0.560	27.690
27	1.392	0.393	49.105	1.391	0.397	14.405	1.377	0.390	27.337
28	1.243	0.923	9.681	1.307	0.884	6.938	1.308	0.864	9.607
29	0.146	0.785	188.665	0.210	0.785	34.868	0.282	0.785	118.620
30	0.001	0.968	54.018	0.017	0.975	18.254	0.001	1.001	35.338
31	0.001	0.885	214.130	0.014	0.905	33.743	0.240	0.785	119.609
32	0.304	1.541	171.444	0.497	1.523	46.128	0.482	1.518	195.447
33	0.036	1.019	36.840	0.001	1.032	19.146	0.001	1.039	37.768
34	1.570	1.401	36.606	1.570	1.402	8.382	1.564	1.395	12.290
35	1.570	0.765	168.059	1.570	0.840	28.236	1.570	0.838	68.381
36	0.001	0.935	105.782	0.001	0.965	22.042	0.001	0.990	55.593
37	0.142	0.561	74.394	0.208	0.560	26.307	0.103	0.478	72.317
38	0.104	0.785	208.096	0.142	0.785	33.595	0.223	0.785	113.323
39	0.001	1.228	86.134	0.001	1.165	24.547	0.001	1.151	70.997
40	1.517	1.032	13.542	1.488	1.056	5.831	1.426	1.352	7.279
41	1.570	1.399	150.382	1.570	1.415	33.612	1.570	1.440	116.251
42	0.469	1.225	104.518	0.738	1.225	39.920	0.850	1.226	175.308
43	0.001	0.675	267.474	0.001	0.627	37.735	0.001	0.591	154.695
44	0.252	0.785	160.979	0.417	0.785	40.637	0.463	0.785	135.393
45	0.235	0.785	170.952	0.386	0.785	40.995	0.457	0.785	138.336
46	1.342	0.590	339.602	0.560	0.785	59.250	0.638	0.785	300.515

次に、算出されたパラメータ η, θ を E_1, E_2, E_3 の誤差の種類毎にヒストグラムで図1から図3に示し、合わせて η, θ の基本統計量として平均、分散、標準偏差を表8に載せる。

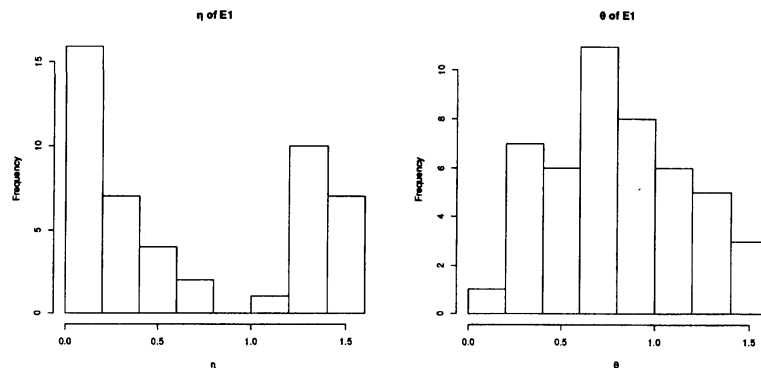


図 1: E_1 の η, θ のヒストグラム

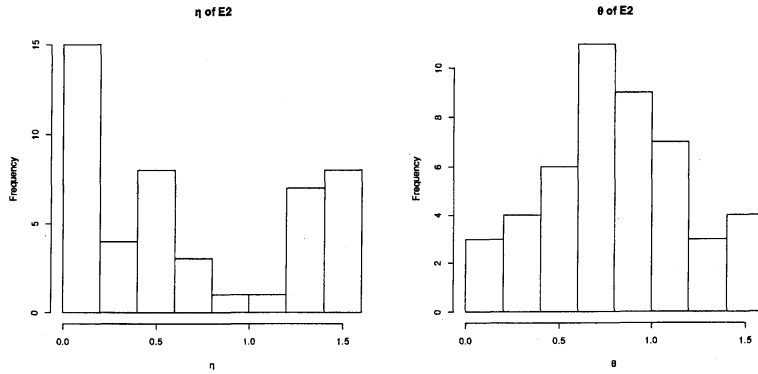


図 2: E_2 の η, θ のヒストグラム

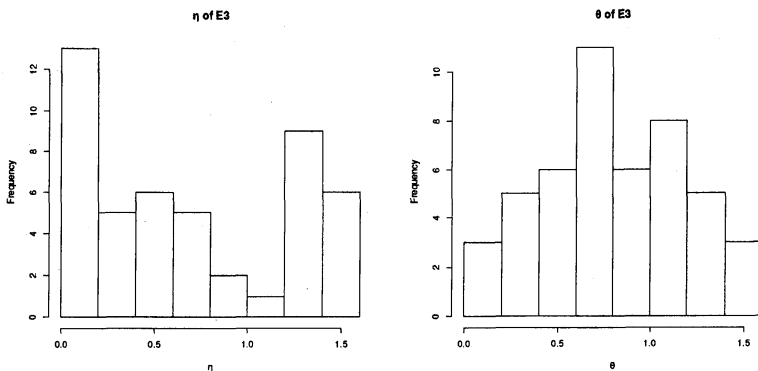


図 3: E_3 の η, θ のヒストグラム

表 8: E_1, E_2, E_3 の最適なパラメータの基本統計量

		平均	分散	標準偏差
E_1	η	0.653	0.372	0.610
	θ	0.816	0.125	0.354
E_2	η	0.651	0.341	0.584
	θ	0.821	0.124	0.353
E_3	η	0.674	0.315	0.561
	θ	0.819	0.140	0.374

これらのヒストグラムを見ると、 E_1, E_2, E_3 で算出されるパラメータの分布の形状に大きな違いは無いことがわかる。全ての誤差の種類で η は範囲の端に峰があり、 θ は 0.6 から 0.7 の間に峰がある。基本統計量を見ても誤差間で算出されるパラメータの大きな違いはない。また、 E_1 の η と θ の相関は -0.06 となり、高い相関は見られなかった。これは E_2, E_3 でも同様の結果が得られた。よって誤差ごとに

算出されるパラメータの類似性から今後は特に何も述べなければ E_1 で算出されたパラメータを分析に使用することにする。 η を見ると端に峰があることから、全体的に歪んでいることがわかる。

ヒストグラムにて η が二峰になっていることから、 η, θ のパラメータセットに対して k-means 法を適応して2つのクラスに分けることにより特性を詳しく見ることにする。 η, θ のパラメータセットに対して k-means 法によって2つのクラスに分けた結果を散布図により図4に示す。 k-means 法は R によって行い、アルゴリズムは Hartigan-Wong を使用した。

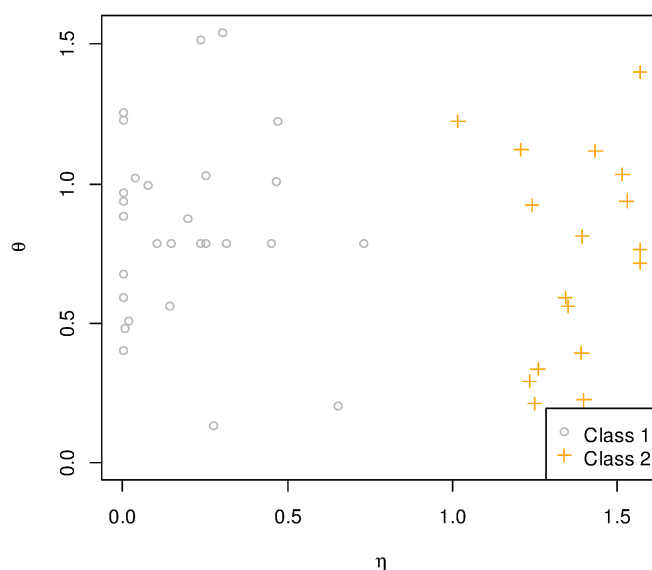


図4: E_1 のパラメータセットに対して k-means 法を適用した結果

k-means 法によるクラスタリングの結果を見ると η の値によって分かれていることがわかる。クラス1の個体数は29、クラス2の個体数は18となった。クラス1とクラス2の η, θ の基本統計量を表9に示す。

表9: クラス毎のパラメータの基本統計量

		平均	分散	標準偏差
Class 1	η	0.201	0.044	0.210
	θ	0.838	0.111	0.333
Class 2	η	1.381	0.025	0.158
	θ	0.780	0.154	0.393

クラス毎の η の平均値から算出される式 (1) の a, b を求めるとクラス 1 の a, b の値はそれぞれ 0.959 と 0.194, クラス 2 の a, b の値は 0.186, 0.971 となり, 約 5 対 1 と 1 対 5 になる. θ に関してはクラス別に大きな差は見られなかった. 表 7 から θ の E_1 の平均を見ると 0.816 となっており, これは $d_{1,1,0}$ の単位円を x 軸の正方向に対して反時計回りに約 45 度回転させることを意味する. よって, 本研究のアンケートから算出された最適距離を実現するパラメータは, 平均的に $d_{1,1,0}$ の単位円を反時計回りに約 45 度回転させ, 回転させたそれぞれの要素の距離の比率を 5 対 1 もしくは 1 対 5 にしていることがわかった.

次に, これまでに算出された基本統計量の区間推定をブートストラップ法によって行う.

4 ブートストラップ法によるパラメータ推定

ここでは, 先の章で求めたパラメータセットを確率変数と見立てて, 基本統計量の区間推定をブートストラップ法によって行う. 本研究では C++ で組んだプログラムによってブートストラップ法を行った. 使用した乱数はセルメヌ・ツイスタ法にて発生させ, 発生させたブートストラップ標本数は 50000 とした. 区間推定は, 対象となるパラメータ η, θ の範囲が決まっていることから, 前提を必要としないパーセントイル法を使用した. E_1 で算出したパラメータセットでブートストラップを行い, 統計量として平均と分散, そして η と θ の共分散と相関に対して算出したブートストラップ標本平均とブートストラップ標本分散, ブートストラップ標本標準偏差と両側 95% ブートストラップ信頼区間の下限と上限を表 10 に載せる. 表中の BS とはブートストラップの略である.

表 10: 各統計量に対する BS 標本平均, BS 標本分散, BS 標本標準偏差とパーセントイル信頼区間

		BS 標本 平均	BS 標本 分散	BS 標本 標準偏差	両側 95%BS 信頼区間	
					下限	上限
平均	η	0.654	0.008	0.088	0.484	0.828
	θ	0.816	0.003	0.051	0.716	0.916
分散	η	0.308	0.001	0.036	0.236	0.377
	θ	0.137	0.001	0.023	0.093	0.184
共分散	$\eta - \theta$	-0.013	0.001	0.031	-0.072	0.048
相関	$\eta - \theta$	-0.062	0.021	0.145	-0.341	0.227

η と θ の共分散と相関を見ると, ブートストラップを行ってもやはり η と θ の関係性が低いことがわかる. 更に, E_1 で算出したパラメータを k-means 法で分けたクラス毎でブートストラップを行った. 表 11, 表 12 にクラス毎のブートストラップの結果を載せる.

表 11: Class1 の各統計量に対する BS 標本平均, BS 標本分散, BS 標本標準偏差とパーセンタイル信頼区間

Class1		BS 標本平均	BS 標本分散	BS 標本標準偏差	両側 95%BS 信頼区間	
					下限	上限
平均	η	0.201	0.001	0.038	0.129	0.278
	θ	0.838	0.004	0.061	0.719	0.957
分散	η	0.043	0.000	0.011	0.022	0.066
	θ	0.107	0.001	0.029	0.055	0.168
共分散	$\eta - \theta$	-0.003	0.000	0.013	-0.031	0.021
相関	$\eta - \theta$	-0.032	0.039	0.197	-0.406	0.361

表 12: Class2 の各統計量に対する BS 標本平均, BS 標本分散, BS 標本標準偏差とパーセンタイル信頼区間

Class2		BS 標本平均	BS 標本分散	BS 標本標準偏差	両側 95%BS 信頼区間	
					下限	上限
平均	η	1.381	0.001	0.036	1.309	1.449
	θ	0.780	0.008	0.090	0.604	0.956
分散	η	0.023	0.000	0.007	0.012	0.039
	θ	0.146	0.001	0.033	0.083	0.213
共分散	$\eta - \theta$	0.014	0.000	0.015	-0.016	0.043
相関	$\eta - \theta$	0.260	0.075	0.275	-0.269	0.787

クラス2の η と θ の相関はこれまでのものより比較的大きいが, 個体数が18と低いいため相関があるとは言い難い。

更にここでは各クラスの η と θ は有意な差があるかをブートストラップ検定によって確認した。クラスの各パラメータ間の平均の差に対してブートストラップを行い, その信頼区間が0を跨いでいるかどうかによって差の検定を行った。差はクラス2の平均からクラス1の平均を引いたものとしている。両側95%ブートストラップ信頼区間に加えて, 両側99%ブートストラップ信頼区間を表13に載せる。

表 13: Class1 と Class2 のパラメータの平均の差に対するパーセンタイル信頼区間

Class1- Class2		両側 95%BS 信頼区間		両側 99%BS 信頼区間	
		下限	上限	下限	上限
平均差	η	1.075	1.281	1.041	1.312
	θ	-0.270	0.156	-0.336	0.225

信頼区間を見ると, η に関しては1%有意で有意な差があり, θ に関しては有意な差が無いことがわかる。よって本研究の事例に対して, 人の感覚は, 歪み方のみに関して2つのグループに分けることが出来ると言える。

5 結論

数量化 III 類を行った分析結果の事例について、人間の感覚に合った距離をよい距離としてアンケートを行い、人の感覚は直角距離で対象間を測ると仮定して最適なパラメータを算出した。更に、そのパラメータを確率変数と見立て、ブートストラップ法によってパラメータの統計量の信頼区間を算出した。

その結果、単位円の縦横への歪みを表すパラメータ η は範囲の両端に峰になり、回転を表す θ は 45 度付近の単峰の形になることが観測された。また、k-means 法により個体はパラメータ η の値のみによって 2 つに分けられることがわかった。またその際、分けられたクラスの a, b の値の平均は約 1:5 と 5:1 となった。これらの結果より、通常よく用いられるユークリッド距離より感覚的な距離に近い他の距離を与えることが出来たといえるだろう。本研究では教育に関する項目に対して計算を行ったが、例えば、商品の特性に対して同様の計算を行うことも可能である。つまり、ある商品群に対して価格や耐久性などの特性値を与え、それらの特性に対して主成分分析や数量化 III 類を適応した結果と、商品群の対比較行列を得る消費者アンケートの結果から、本研究と同様の η と θ の分布を得ることが出来る。その分布から消費者がどのような感覚で商品の違いを認識しているか、また、その分布に対してクラスター分析を行うことによって、商品群に対する消費者の感覚はどのようなグループで分けることが出来るかを知ることが可能になることを本研究は示唆する。

今後の課題としては i) 他の多くの事例に対して同様の計算・分析を行うこと、ii) 多くの人の各核的な距離に合った距離選択はどのようになるか、多くの人の感覚的な距離の対比較データを集めることによってさらに実験を進めること、iii) 他の距離関数も考慮すること、等が考えられる。

参考文献

- [1] 青森県総合社会教育センター、家庭の教育力に関する調査報告書, 2002.
(http://www.alis.pref.aomori.lg.jp/research/research_h13katei.html)
- [2] M. R. Chernick, *Bootstrap Methods: A Practitioner's Guide*, Wiley, USA, 1999.
- [3] M. Kon, *Public opinion survey on home education: application of location problems with rectilinear norm*, *Scientiae Mathematicae Japonicae*, **58**, 2003, 99-111.

- [4] R. F. Love and P. D. Dowling, *Optimal weighted l_p norm parameters for facilities layout distance characterizations*, Management Science, **31**, 1985, 200-206.
- [5] R. F. Love and J. G. Morris, *Modelling inter-city road distances by mathematical functions*, Operational Research Quarterly, **23**, 1972, 61-71.