

1 II. 数量的解析の試行。

2 II-1. 材料 (Excel dataset)

3 ここで分析対象にするのは、Berg-Schlosser D. and De Meur(1994)の、Conditions of  
 4 Democracy in Inter war Europe. A Boolean test of major hypotheses. 「戦間期に送るヨーロ  
 5 ッパの民主主義の生存条件。ブール演算による主要な仮説の検証」において、分析対象に  
 6 なったデータセットである。この本では、Benoit Rihoux and Charles Ragin (2009)  
 7 CONFIGURATIONAL COMPARATIVE METHOD : Qualitative Comparative  
 8 Analysis(QAC) and Related Technique で csQCA, mvQCA, fsQCA を紹介する際に用いた  
 9 データに、政権の安定性という項目を加えている。Berg-Schlosser, De Meur(1994)が行っ  
 10 た分析の目的は、Lipset(1960)が唱えた、「近代化が民主主義を発展させる」という仮説の  
 11 検証である。本解説でも、この仮説の検証を意識して、分析方法の有効性を論ずる。

12 表1に分析対象とするデータセットを示した。分析対象は、第一次世界大戦の終了から、  
 13 第二次世界大戦の開始時で(1919-1939)、ヨーロッパに存在した、オーストリア、ベルギー、  
 14 チェコスロバキア、エストニア、フィンランド、フランス、ドイツ、ギリシャ、ハンガリー、  
 15 アイルランド、イタリア、オランダ、ポーランド、ポルトガル、ルーマニア、スペイン、ス  
 16 エーデン、大英帝国の18か国であり、データ項目は、GNP：1930年時点における国民一人  
 17 当たりの GNP、都市化:人口2万人以上の都市に住む都市人口の全人口に対する百分率、  
 18 教育：識字率%、工業人口:工場労働者の全労働人口に対する百分率、安定性（非安定

19 表1. 分析対象のデータセット

		GNP	都市化	教育	工業人口	安定性	保全度
国	事例 Id	A	B	C	D	E	R
オーストリア	AUT	720	33.4	98.0	33.4	10	-9
ベルギー	BEL	1098	60.5	94.4	48.9	4	10
チェコスロバキ	CZE	586	69.0	95.9	37.4	6	7
エストニア	EST	468	28.5	95.0	14.0	6	-6
フィンランド	FIN	590	22.0	99.1	22.0	9	4
フランス	FRA	983	21.2	96.2	34.8	5	10
ドイツ	GER	795	56.5	98.0	40.4	11	-9
ギリシャ	GRC	390	31.1	59.2	28.1	10	-8
ハンガリー	HUN	424	36.3	85.0	21.6	13	-1
アイルランド	IRL	662	25.0	95.0	14.5	5	8
イタリア	ITA	517	31.4	72.1	29.6	9	-9
オランダ	NLD	1008	78.8	99.9	39.3	2	10
ポーランド	POL	350	37.0	76.9	11.2	21	-6
ポルトガル	PRT	320	15.3	38.0	23.1	19	-9
ルーマニア	ROU	331	21.9	61.8	12.2	7	-4
スペイン	ESP	367	43.0	55.6	25.5	12	-8
スウェーデン	SWE	897	34.0	99.9	32.3	6	10
大英帝国	UK	1034	74.0	99.9	49.9	4	10

表2.標準化したデータリスト (Excel standardize)

Case Id	A	B	C	D	E	R
AUT	0.30	-0.35	0.73	0.40	-0.24	-1.12
BEL	1.75	1.09	0.54	1.75	0.98	1.24
CZE	-0.21	1.54	0.62	0.75	0.58	0.87
EST	-0.66	-0.61	0.57	-1.28	0.58	-0.74
FIN	-0.20	-0.95	0.79	-0.59	-0.03	0.50
FRA	1.31	-0.99	0.63	0.52	0.78	1.24
GER	0.59	0.88	0.73	1.01	-0.44	-1.12
GRC	-0.96	-0.47	-1.36	-0.06	-0.24	-0.99
HUN	-0.83	-0.19	0.03	-0.62	-0.85	-0.12
IRL	0.08	-0.79	0.57	-1.24	0.78	0.99
ITA	-0.47	-0.45	-0.66	0.07	-0.03	-1.12
NLD	1.40	2.06	0.83	0.91	1.39	1.24
POL	-1.11	-0.16	-0.40	-1.53	-2.47	-0.74
PRT	-1.23	-1.30	-2.49	-0.49	-2.06	-1.12
ROU	-1.19	-0.95	-1.22	-1.44	0.37	-0.50
ESP	-1.05	0.16	-1.55	-0.29	-0.64	-0.99
SWE	0.98	-0.31	0.83	0.30	0.58	1.24
UK	1.50	1.80	0.83	1.83	-0.98	1.24

22

23 性)：戦間期20年間の政権交代の回数、それに、結果(被説明変数)としての、民主主義  
 24 の保全程度の評価値(-10から10)(Berg-Schlosser and Mitchell(200,2003))である。

25 分析データは、個々のデータ項目が結果に及ぼす影響がわかりやすいように、平均値を  
 26 差し引いた後に、データの標準偏差で割って標準化した。Eの政治的安定性については、  
 27 政府の符号を入れ変えた(Excel standardize)。このデータを用いて、相関分析、多次元尺  
 28 度構成法、クラスター分析、主成分分析、因子分析、回帰分析を行う。

29

30

31 I I- 2. 総当たり相関分析 (R:line20-25)

32

33 表3に総当たりの相関分析の結果を示した。計算はR (line20-25)で行った。結果(目的  
 34 変数 R)との相関係数はいずれのデータ項目でも無視できない値である。総当たり相関の場  
 35 合、すべての組み合わせ(6 x 5)で、相関係数を計算しているの、確率論的に有意性を  
 36 検定する場合には、繰り返しによる補正を行う必要がある(例えば、Bonferroni補正)。そ  
 37 の結果として、B、DとRの相関の有意性は確率的に否定されるが、私たちは、この結果  
 38 を何かの予想に使おうとしているわけではない。この場合、統計的な検定にはあまり意味  
 39 がない。むしろ、B、DとRの間に無視できない程度の大きさで相関があったのだが、そ  
 40 の相関はA、C、EとRの相関に比べて低い値であった、という事実の方が重要かもしれ  
 41 ない。このことを、記憶したうえで、次の解析に進む。

42

43

表3. 総当たり相関分析)

総当たり相関表							危険率						
	A	B	C	D	E	R		A	B	C	D	E	R
A	1.00	0.56	0.73	0.80	0.69	0.74		0.00	0.11	0.01	0.00	0.02	0.01
B	0.56	1.00	0.44	0.72	0.41	0.41		0.02	0.00	0.35	0.01	0.35	0.35
C	0.73	0.44	1.00	0.41	0.62	0.63		0.00	0.07	0.00	0.35	0.05	0.05
D	0.80	0.72	0.41	1.00	0.47	0.44		0.00	0.00	0.09	0.00	0.28	0.35
E	0.69	0.41	0.62	0.47	1.00	0.69		0.00	0.10	0.01	0.05	0.00	0.02
R	0.74	0.41	0.63	0.44	0.69	1.00		0.00	0.09	0.01	0.07	0.00	0.00

44

45 危険率の上三角行列は繰り返しを考慮した確率。下三角行列は繰り返しを考慮していな  
 46 い。

47

48 II - 3 . 距離行列(Excel dist)とデータの空間的位置関係 (MD S :R line65-74)

49

50 このデータを直接、回帰分析などに持っていても、多重共線性の問題があり、うまく  
51 解析が進まないだろう。こういう場合、サンプル（この場合は国）の類似性あるいは違い  
52 によって、タイプ分けすると、議論を単純化するのに役立つだろう。タイプ分けに使える  
53 方法は、様々あるが。考えやすいのは、それぞれの国ごとの違いを距離で表して、距離行  
54 列を作るというやり方である。その場合、問題になるのは、項目間の分散（表現している  
55 データの尺度の違い）の違いと相関の存在である。私たちは、すでに標準化したデータを使  
56 っているから、項目間の分散の違を補正する必要はないが、項目間に相関の影響は考慮  
57 しなければならない。これを無視すると、或る方向への変化を、重複して評価してしまう  
58 ことになる。そのため、データ間の相関を0にするような、変換を行わなければならない。  
59 線形代数学の用語では、これを相関行列の対角化という。対角化すると相関行列の非対角  
60 因子（相関）が0になるので、変換された軸はすべて直交していることになる。データの  
61 行列を直交する行列に変換して求めた距離を、マハラノビス距離という（図1）。中学校で  
62 は、直角三角形の斜辺の2乗は他の2辺の二乗の和に等しいと習った。多次元空間に拡張  
63 すれば、多次元空間における2点間の距離の2乗は、

64 
$$d^2 = \mathbf{u}\mathbf{u}^t = u_1^2 + u_2^2 + \dots$$

65 距離は

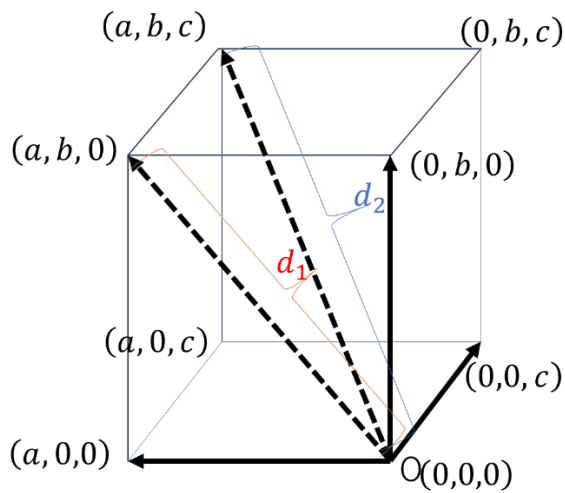
66 
$$d = \sqrt{\mathbf{u}\mathbf{u}^t}$$
  
67 
$$\mathbf{u}$$
 は項目ごとの差のベクトル( $u_1 \ u_2 \ \dots$ )

68 となるのだが、何しろ、あのピタゴラスの定理だから、この計算は、各項目が相関を持た  
69 ずに直交していることを、暗黙の前提としている。これをユークリッド距離と云う。実  
70 際のデータは項目間に相関があるから、項目間軸を変換して、直交化してから距離を計算  
71 しなければならない。マハラノビスの距離の式は

72 
$$md = \sqrt{\mathbf{u}\mathbf{r}^{-1}\mathbf{u}^t}$$
  
73 
$$\mathbf{r}$$
: 相関行列

74 
$$\mathbf{u}$$
: 観測項目ごとの差の行列  $\mathbf{u} = (u_1 \ u_2 \ \dots)$

75 である。マハラノビスの距離の公式をどのように導くかは図1に示した。要約すると、ユ  
76 ークリッド距離の二乗を求める式は、項目ごとの差の行列にその転置行列を掛けるという



Pythagorean theorem

$$d_1^2 = a^2 + b^2$$

3 dimensional expansion

$$d_2^2 = a^2 + b^2 + c^2$$

Matrix operation

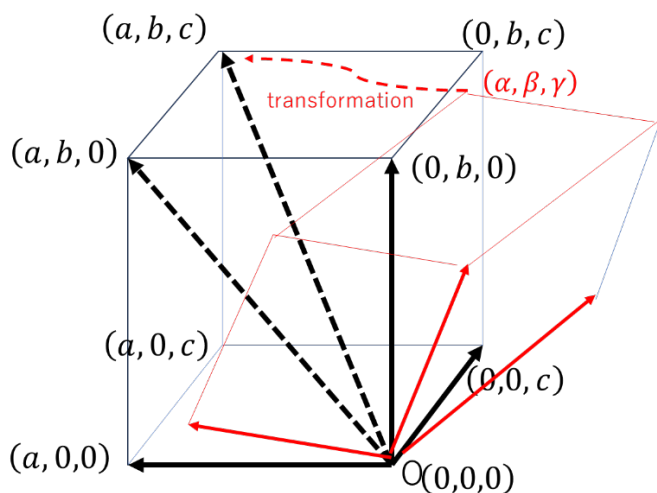
$$d_2^2 = (a \ b \ c) \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Euclid distance

$$d_2 = \sqrt{\mathbf{u}\mathbf{u}^t}$$

$\mathbf{u}^t$ : transpose of  $\mathbf{u}$

77



Mahalanobis distance

$$d_2^2 = (a \ b \ c) \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$= (\alpha \ \beta \ \gamma) \mathbf{r}^{-\frac{1}{2}} \left( (\alpha \ \beta \ \gamma) \mathbf{r}^{-\frac{1}{2}} \right)^t$$

$$= (\alpha \ \beta \ \gamma) \mathbf{r}^{-\frac{1}{2}} \mathbf{r}^{-\frac{1}{2}t} (\alpha \ \beta \ \gamma)^t$$

$$= (\alpha \ \beta \ \gamma) \mathbf{r}^{-\frac{1}{2}} \mathbf{r}^{-\frac{1}{2}t} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$$

$$= (\alpha \ \beta \ \gamma) \mathbf{r}^{-1} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$$

$$d_2 = \sqrt{\mathbf{u}\mathbf{r}^{-1}\mathbf{u}^t}$$

Where

$\mathbf{r}$  is correlation matrix

78

79 図1. マハラノビスの距離の公式の導出

80

81 式だが、その間に相関行列の逆行列を挟んで行列演算をすれば、マハラノビス距離の二乗  
 82 になる。式としては簡単である。総当たりで距離を計算するのは、ユークリッド距離でさ  
 83 えめんどうだ。Rで計算しようと思って、dist という距離行列を作る関数を調べたら、距  
 84 離の指定の所に、何故か、Mahalanobis 距離がない。マハラノビス距離を計算する、  
 85 mahalnobis という関数もあるのだが、こちらは、総当たりに距離行列を作ってくれな  
 86 い。国が18あると、18×17回も、計算を繰り返さなければならない。そんなことを

87 やるぐらいならば、Excel でやった方が簡単かもしれない。ということで、Excel で距離行  
88 列を作った。データの行列をまとめて計算したので。計算回数は18回で済んだ。その過  
89 程を Excel dist に残した。Mahalanobisi 距離の計算のテンプレートとして使える。  
90 Mahalanobisi 距離の理解の役立つ。

91 距離行列を表4に示した。距離行列を見ていても、どこをどう見たらわからないが、要  
92 約のように、距離の近い組み合わせ、距離の遠い組み合わせについて整理すると、結果が  
93 見やすい。こういう時は Excel の「並べ替え」の機能が役立つ。条件が正しく選ばれていて、  
94 その条件が同じならば、結果が同じになるだろうと考えられるが、条件が似たような組み  
95 合わせ、具体的には、距離近さが、10番目まで(全体の3.3%)の組み合わせの内、同じ結  
96 果になったのは、**イタリアーギリシャ、ドイツーオーストリア、スペインーギリシャ、ハン**  
97 **ガリーーオーストリアハンガリーードイツ、ベルギーー大英帝国、スウェーデンーフラン**  
98 **スの7つの組み合わせである。結果が異なったのは、**オーストリアーフィンランド、ハン****  
99 **ガリーーフィンランド、アイルランドーエストニア**の組み合わせである。10の組み合わ  
100 せのうちで、7つの組み合わせで、同じということは、民主主義の維持・崩壊という結果  
101 を招く条件として選ばれたA、B、C、D、Eの条件に、全体としてある程度の妥当性が  
102 あることを示していると考えられる。しかし、さらによく見ると、民主主義が崩壊した国  
103 の中で、強い類似性もった、ドイツーオーストリアーハンガリーという一つの塊が見えて  
104 くる、これらの国と、フィンランドは距離的に近い、しかし、結果は異なっている。つま  
105 り、フィンランドという国は、条件的には、民主主義が崩壊する条件がそろっていたにも  
106 かかわらず、何かの要因によって、民主主義が守られたのだと考えられる。フィンランド  
107 はソ連とドイツに挟まれて、侵略を受けながらも、様々な戦略的・外交的手段を使って、  
108 独立を守り抜いた。そういうことも、民主主義の維持と関係しているかもしれない。同様  
109 に異なる結果となったアイルランドとエストニアは、ともに戦間期に他国の支配を離れて、  
110 独立したという共通点を持っている。そのため、都市化や工業化などは共に遅れていたと  
111 考えられる。数値的に見ると政権も比較的安定していたと言えるが、独立後、間もないた  
112 めに様々な不安定要因もあったであろう。その中であって、何が、アイルランドの民主主  
113 義を守る要因となったのか、比較研究が必要であろう。条件の距離が遠く離れた国の組み  
114 合わせを見ると、距離が遠い10の組み合わせの内、5つの国で同じ結果となっている。  
115 その内、4つは民主主義が崩壊した国々の組み合わせである。条件が異なる10の組み合  
116 わせのうち、5つの条件のすべての組み合わせが同じような位置になくても、民主主義が  
117 崩壊するのであれば、5つ全ての条件がそろわなくても、もっと少数の、あるいはデータ  
118 として示されていない条件の違いによって、結果が導かれた可能性を示唆している。特に、  
119 5つの内4つの組み合わせが、民主主義の崩壊という結果になっていることから、民主主  
120 義の崩壊は、多くの条件の組み合わせではなくて、いくつかの特定の条件によって崩壊し  
121 た可能性を示している。こういう分析を MDSO/MSDO とする。MDSO/MSDO とは、  
122 Most different, Similar outcome/Most similar, different outcome の略である。

123

表4. マハラノビス距離による距離行列 (Excel Dist)

	AUT	BEL	CZE	EST	FIN	FRA	GER	GRC	HUN	IRL	ITA	NLD	POL	PRT	ROU	ESP	SWE
BEL	2.86																
CZE	3.07	3.77															
EST	2.85	3.83	2.99														
FIN	1.26	3.44	3.23	1.88													
FRA	2.50	2.37	4.78	3.54	2.62												
GER	1.52	2.47	2.69	3.36	2.42	3.28											
GRC	2.78	3.05	2.88	3.01	2.93	3.42	3.04										
HUN	1.75	3.34	2.44	2.09	1.66	3.53	1.77	2.51									
IRL	3.39	3.38	4.27	1.80	2.62	2.86	3.77	3.61	2.97								
ITA	2.01	2.61	2.71	2.56	2.19	2.76	2.46	0.83	2.02	3.14							
NLD	4.32	2.57	3.86	3.67	4.35	4.18	3.59	4.09	3.78	3.10	3.85						
POL	4.09	4.61	4.95	4.21	4.07	4.96	3.48	4.83	3.03	3.98	4.51	4.28					
PRT	3.65	3.51	4.98	4.59	4.05	3.64	3.63	2.74	3.48	4.34	2.83	4.79	4.03				
ROU	3.94	3.70	3.91	2.29	3.39	3.72	4.24	2.39	3.18	2.25	2.47	3.53	4.70	3.68			
ESP	3.34	2.93	3.02	3.13	3.47	3.91	2.99	1.56	2.44	3.40	1.86	3.18	3.88	2.53	2.22		
SWE	1.85	1.94	3.79	2.60	1.88	1.30	2.34	3.00	2.47	2.10	2.27	3.23	3.99	3.53	3.18	3.18	
UK	2.76	1.27	2.73	3.56	3.35	3.25	1.95	2.97	2.90	3.60	2.58	2.27	4.44	3.99	3.80	2.72	2.41

124

125

要約

近い順				遠い順		
起点	終点	距離		起点	終点	距離
ITA	GRC	0.83		PRT	CZE	4.98
AUT	FIN	1.26		POL	FRA	4.96
BEL	UK	1.27		POL	CZE	4.95
SWE	FRA	1.30		POL	GRC	4.83
GER	AUT	1.52		PRT	NLD	4.79
ESP	GRC	1.56		FRA	CZE	4.78
HUN	FIN	1.66		ROU	POL	4.70
HUN	AUT	1.75		POL	BEL	4.61
HUN	GER	1.77		PRT	EST	4.59
IRL	EST	1.80		POL	ITA	4.51

126

127 MDSO/MSDO はアンケート調査の分析などに用いられ、距離としては、情報理論で用い  
 128 られるハミング距離などが使われ、QCA の手法の1つとされる。ここでは、  
 129 MDSO/MSDO を理解するために、あえて、連続的数値な数値データでも、似たようなこ  
 130 とが可能なることを示してみた。

131 次に、距離行列を使って、各国が置かれていた条件の違いや類似性を全体として認識す  
 132 るために、5次元の距離行列を2次元平面に落として視覚化した(図2)。用いたのは多次元  
 133 尺度構成法(MDS: Multi-dimensional scaling method)である(R script line65-74)

134 MDSは類似度や距離などを使って、多次元の調査対象間の位置関係を、2次元あるいは3  
 135 次元空間の位置で表現する手法で、アンケート調査や生態学的な調査の解析などで良く用  
 136 いられる。MDSの縦横の軸には意味がないので、MDSを見る時には、自分で図を回転さ  
 137 せながら見るのだが、反時計回りに少し回転させてみると、民主主義崩壊国と民主主義維  
 138 持国が、上下に偏っているように見える、上側には、ポーランド、ルーマニア、ハン  
 139 ガリー、スペイン、イタリア、ギリシャ、ポルトガルの民主主義崩壊国が位置し、下側に、

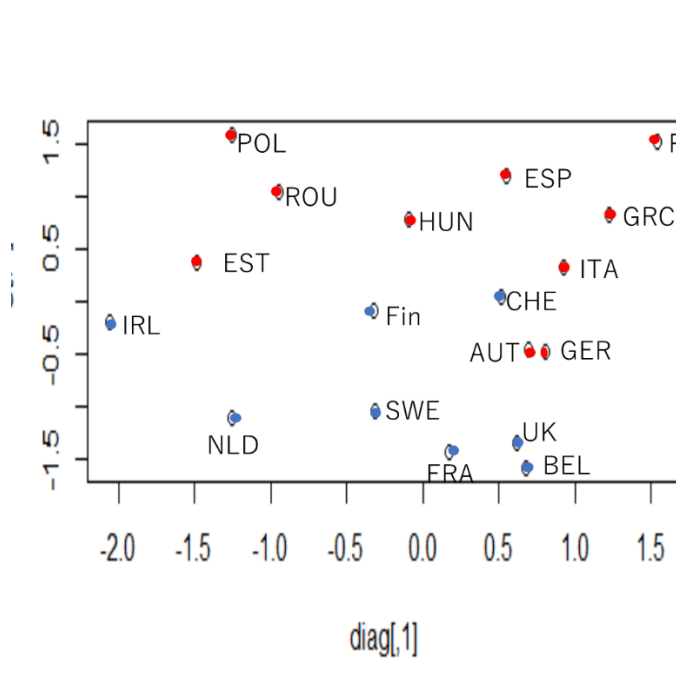


図 2. MDS で描いた各国の関係 (R script line71-74)

140

141

142

143 オランダ、スウェーデン、フランス、連合王国、ベルギーの民主主義維持国が位置してい  
 144 る。その中間帯には、アイルランド、エストニア、フィンランド、チェコ、ドイツ、オー  
 145 ストリアが位置している。この中には、民主主義維持国と民主主義崩壊国が3か国ずつ  
 146 含まれる。中間帯に位置する国を除いて、それらの国々について、民主主義の崩壊要因、  
 147 維持要因を分析するのは容易で、おそらく、Lipset(1960)の仮説の妥当性が検証されるだ  
 148 ろう。ここで問題になるのは、何故、中間帯に位置する3か国において民主主義が維持さ  
 149 れ、他の3か国で民主主義が崩壊したのかだろう。解説者は社会学者でも政治学者でもな  
 150 いから、その分析はできない。適切な社会学者であれば、この6か国間の精細な比較から、  
 151 何らかの結論を導き出せるのではないかと思う。豊かさ、教育、政治の安定性などとは、  
 152 違う次元で、チェコスロバキア、フィンランドと、ドイツ、オーストリアの違いを見つけ  
 153 て、それらを加えて、解析をやり直すというのが、あるべき、解析の方向である。

154



155 II-4. 主成分分析 (R line32-62)

156

157 マハラノビスの距離も直交化の一つだが、距離行列の計算は面倒だし、その後、クラス  
 158 ター分析などに持って行くのも、あまりうまくいかなかった（解説者がRの script の書き  
 159 方がわからないだけかもしれない。）ので、手っ取り早く直交化する方法として、主成分分  
 160 析を行う。主成分分析は、分散共分散行列、相関行列の対角化であり、類似性に基いた、  
 161 クラスター分け等に使える。

162 主成分分析の結果の要約を表5に示した。標準偏差で標準化してあるので、固有値の合  
 163 計(全分散)は5である。分散1の項目が5つあるのだから全分散は5に決まっている。そ  
 164 の内、66.8%の分散を第一主成分が占め、第二主成分が、17.4%を占めている。累積では、  
 165 この2つの主成分が全体の8割以上を占めていて、他の主成分はマイナーな主成分である。  
 166 PC1について見ると、各項目の負荷量は、すべてマイナスで矢印の長さも似たような値で  
 167 ある。すべての測定項目の変動が、何か特定の方向の変動と結びついている。表6に第一  
 168 主成分と第二主成分について、負荷量の大きな項目名を用いて、主成分が表現する内容の  
 169 考察を示した。第一主成分について見ると、貧しい、都市化されていない、教育レベルが  
 170 低い、工業化していない。政治の不安定が並ぶ。相関分析では、A（豊かさ一貧しいと正  
 171 負が反対）は他の4つの項目と高い相関を示していた。したがって、第一主成分を、貧し  
 172 さにかかわる主成分と考えることもできるが、第五主成分は明らかに、豊かさにかかわる  
 173 主成分である。分散が小さいので、第5主成分を無視しても良いのだが。第5主成分が豊  
 174 かさだとすれば、第1主成分はもっと包括的な社会の構造を表しているのだと考えられる。  
 175 そこで、Lipset(1960)の仮説を考慮して、第1主成分に後進性（反近代性）という名前を

176

177

表5 主成分分析の結果の要約 (R script 26-48)

	主成分				
x	PC1	PC2	PC3	PC4	PC5
固有値	3.338	0.871	0.396	0.331	0.064
割合	0.668	0.174	0.079	0.066	0.013
累積割合	0.668	0.842	0.921	0.987	1
	負荷量				
	PC1	PC2	PC3	PC4	PC5
A	-0.51	0.141	-0.17	-0.42	0.715
B	-0.40	-0.63	0.127	0.617	0.233
C	-0.44	0.406	-0.63	0.381	-0.32
D	-0.47	-0.44	0.006	-0.53	-0.56
E	-0.41	0.483	0.749	0.125	-0.14

178

表 6. 第一主成分と第二主成分についての考察

PC1

- A(-0.51): 金持ちではない → 貧乏
- B(-0.40) 都市化されていない
- C(-0.44) 低い識字率 → 低い教育レベル
- D(-0.47) 工業化していない
- E(-0.41) 不安定な政治

後進性 ?

PC2

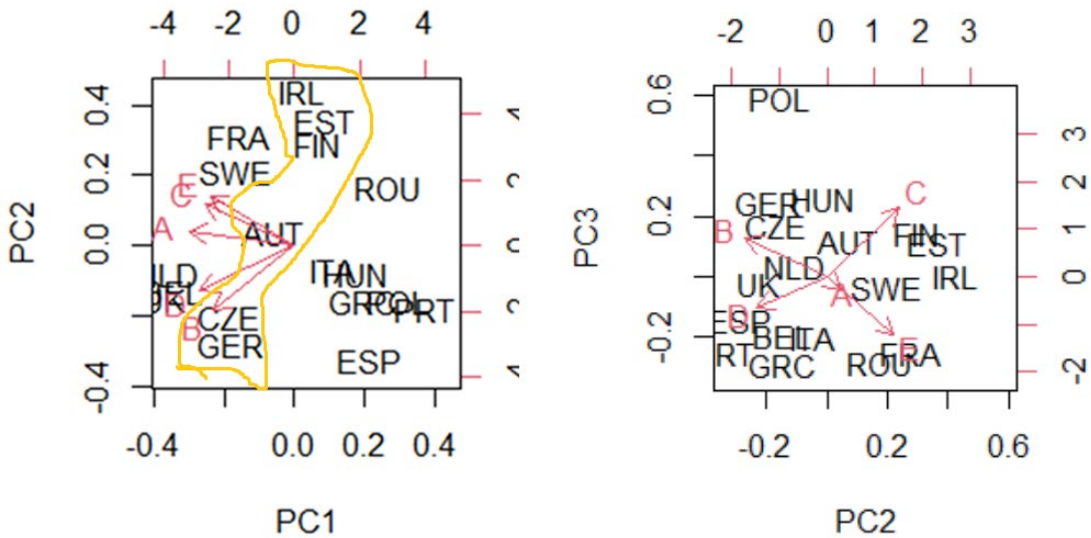
- B(-0.63) 都市化されていない
- C(0.41) 高識字率 → 高い教育レベル
- D(-0.44) 工業化していない
- E(-0.48) 安定した政治

先進的な農業国 ?

180

181 付けた。第二主成分は、第一主成分と同様に、B の都市化と工業化がマイナスであるが、  
 182 C の識字率がプラスの方向で、政治も安定している。このことから考えられる国のイメージ  
 183 は豊かで安定した農業国である。そこで、第 2 主成分は、農業的とした。第 3 主成分は、  
 184 E の負荷量が大きく、政治的な安定性を示している主成分と考えられる。第 4 主成分は都  
 185 市化である。

186 図 3 は、これらの主成分得点を使って描いた、各国の散布図である。この図の中に赤い  
 187 矢印で、各データ項目が示すベクトルの方向を示した。PC1-PC2 のプロットの中で、A の  
 188 国の豊かさは、PC1 の正の方向とほぼ 180 度反対の方向を向いている。またその他の項目  
 189 も PC1 と反対の方向を向いている。PC1-PC2 の散布図に、アイルランド、フィンランド、  
 190 ド、チェコスロバキア、エストニア、オーストリア、ドイツのグループを黄色い線で囲んだ。  
 191 このグループから右側では、すべての国が、民主主義崩壊国であり、左側では、すべて民  
 192 主主義維持国である。ここで、この中間帯が、時計方向にやや傾いて、斜めになっている



193

194

図 3. 第 1 主成分、第 2 主成分、第 3 主成分で描いた、各国の散布図

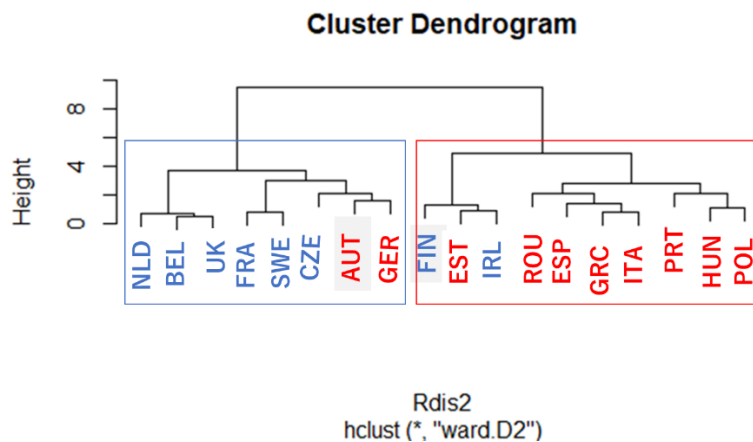


図 4. 主成分を使ったクラスター分析 (デンドログラム:R lin57-62)

195  
196  
197  
198 というのが、後の解析で重要な意味をもつことになる。いずれにしても、この解析からも。  
199 豊かさを含む何か (近代化?) が、民主主義維持にかかわる要因であるという、Lipset の  
200 仮説が部分的には支持される。主成分分析では、このように、第 1 主成分に、データがセ  
201 ットが持つ多面的な分布特性を包括的に表す、よくわからない成分が抽出されることが多  
202 い。これをどのように解釈するかは、研究者をしばしば悩ませる。現時点で解説者が言え  
203 ることは、「第一主成分は、豊かさや教育程度政治の安定等々に包括的にかかわる何かだ。」  
204 ということだけである。Lipset(1960)に倣って第 1 主成分を反近代化と表現すべきか、他  
205 の言葉で表現すべきかは、過去の研究事例、経験等々、何らかの根拠をもって、分析者が  
206 判断すべき事項である。いずれにしても、第 1 主成分と第 2 主成分で、**アイルランド**、**フ**  
207 **ィンランド**、**チェコスロバキア**、**エストニア**、**オーストリア**、**ドイツ**の違いをもたらした  
208 要因を明らかにすることはできない。図 4 に主成分得点を使ったクラスター分析の結果を  
209 示す。主成分は直交しているので、距離はユークリッド距離、クラスタリングの方法は  
210 Ward 法を使った。赤い線で囲った主として民主主義崩壊国で構成される、右側の大きなク  
211 ラスターの中の一つのサブクラスターを、**フィンランド**と**アイルランド**が**エストニア**とと  
212 もに構成している。おなじように、青い線で囲った、主として民主主義維持国で構成され  
213 る、左側の大きなクラスターの中のサブクラスターを、**オーストリア**と**ドイツ**が**チェコ**と  
214 ともに構成している。このデンドログラムが表現しているのは、民主主義崩壊の可能性が  
215 高かった国々の中で、**フィンランド**、**アイルランド**を特殊とし、民主主義維持の可能性が  
216 あった国の中で、**ドイツ**、**オーストリア**を特殊とする視点である。

217

218 II-5. 回帰分析(R line76-121)

219

220 以上のようなデータ分布の特性を理解した上で、回帰分析を行う。回帰分析は、目的変  
 221 数を R とする。説明変数を標準化したデータ項目 A、B、C、D、E とした回帰分析（回帰  
 222 分析 1）と、説明変数を各主成分の主成分得点とした（回帰分析 2）を行う。回帰分析 1  
 223 が対象とするデータリストは表 2 のデータリストである。回帰分析 2 で分析するデータリ  
 224 ストを表 7 に示した。

225 回帰分析 1 の結果は以下の通りであった。すでに標準化されたデータを用いているの  
 226 で、定数項（切片 = 0）はない。各説明変数の説明力を比較するために分散比を示した。

227

		単回帰		
		PR(>t)		分散比
230	R=0.7399A	0.000448**		0.547
231	R=0.4072B	0.0935		0.166
232	R=0.6269C	0.00536**		0.393
233	R=0.4360D	0.0765		0.190
234	R=0.7399E	0.00158**		0.474

235

		重回帰			分散比
237		R=0.4240A+0.1077C+0.371E			0.610
238	PR(>t)	0.128	0.672	0.200	

239

240

241

表 7. 回帰分析 2 のデータリスト

ID	主成分得点					標準化後のR
	PC1	PC2	PC3	PC4	PC5	
AUT	-0.46959	0.147189	-0.91418	-0.42496	-0.24998	-1.11631
BEL	-2.64316	-0.50358	0.387021	-0.27646	0.187727	1.240347
CZE	-1.46529	-0.80074	0.326466	0.887345	-0.58806	0.868243
EST	0.686033	1.414189	0.152967	0.884029	-0.22566	-0.74421
FIN	0.515894	1.14744	-0.72665	0.14239	-0.20949	0.496139
FRA	-1.22524	1.225289	-0.09471	-1.1583	-0.00669	1.240347
GER	-1.38023	-1.1128	-0.9795	-0.01534	0.017861	-1.11631
GRC	1.562003	-0.62098	0.711752	-0.47955	-0.23846	-0.99228
HUN	1.369426	-0.30518	-0.69783	0.487664	-0.04962	-0.12403
IRL	0.214302	1.752321	0.153952	0.419581	0.298637	0.992278
ITA	0.881597	-0.28057	0.502764	-0.5062	-0.21354	-1.11631
NLD	-2.77554	-0.30509	0.472418	0.407505	0.303656	1.240347
POL	2.291442	-0.63537	-0.81251	0.657366	0.446106	-0.74421
PRT	2.87268	-0.72202	-0.20397	-0.84476	0.03812	-1.11631
ROU	2.055549	0.665163	1.245056	0.103768	0.112798	-0.49614
ESP	1.663234	-1.2918	0.49379	0.0945	0.16194	-0.99228
SWE	-1.27245	0.836212	-0.26645	-0.44917	0.117729	1.240347
UK	-2.88066	-0.60968	0.249597	0.070591	0.096918	1.240347

242

243 分散比は、回帰式で説明可能な分散の全分散に対する比である。Excel を使って、1-誤差  
 244 分散/全分散の式で計算した。この計算過程は、Excel シート V ratio に残した。最も係数の  
 245 絶対値が大きいのが、Aの豊かさであり、誤差分散を含む全分散の 50%以上説明し統計的  
 246 に極めて有意である。次に係数の絶対値が大きいのは、E の政治の安定性であり、45%以  
 247 上を説明し統計的に有意である。次に係数が大きいのが、Cの識字率であり、全体の 40%  
 248 近くを説明する。これらをたすと、3つの項目で、全体の 140%を説明することになる。  
 249 これは、因子間に相関があるために、説明力を重複してカウントしているためである。こ  
 250 れら A、C、E を説明変数として、重回帰分析を行うと、A の係数だけが有意になり、他  
 251 は有意でない。この分散比は 0.610 であり、この回帰式の持つ説明力は、全体の分散の  
 252 60%である。単回帰で計算した説明力の和の 140%に比べて、極めて小さい。重回帰によ  
 253 って、相関による重複したカウントを除くと 80%も小さくなる。このことは、単回帰では、  
 254 2重3重に重複して説明力が計算されていたことを示している。

255 重複した説明力のダブルカウントを防ぐために、主成分分析によって直交化した主成分  
 256 を説明変数とした回帰分析 2 (R line78-100)の結果を示す。

257

258

単回帰

		PR(>t)	分散比
259			
260	$R = -0.4040PC1$	0.000805***	0.514
261	$R = 0.3806PC2$	0.161	0.119
262	$R = 0.2297PC3$	0.578	0.018
263	$R = 0.1131PC4$	0.803	0.004
264	$R = 0.7964PC5$	0.436	0.038

265

重回帰

266	$R = -0.404PC1 + 0.3906PC2 + 0.7963PC5$		0.672
267	PR(>t)	0.00035 *** 0.04065* 0.22164	

268

269 第1主成分の係数は負の値である。これは第一主成分が後進性あるいは貧しさというネ  
 270 ガティブな傾向を正としているためである。この式単独でも、半分以上(0.514)の分散を説  
 271 明できる。第二主成分は、それよりも説明力が小さく、説明できる分散比は 0.119 である。  
 272 それ以下の主成分の説明力はさらに小さくなるが、第5主成分は、分散が小さいにもかか  
 273 わらず。第3、第4主成分よりも説明力が高い。第1主成分、第2主成分、第5主成分を  
 274 使って重回帰分析をすると、第1主成分は極めて有意であり、第2主成分も有意になる。  
 275 この分析では、元々サンプルサイズが小さいので、自由度が不足して、有意でないと判定  
 276 される。だが、すでに述べたように、この解説では統計的優位性を論ずることに意味がな  
 277 い。第5主成分を加えた式の、分散比は 0.672 で、データ項目の A、C、E の3つを説明変  
 278 数とした式よりも説明力が大きくなる。この分散比は、単回帰分析の分散比の和

279 (0.514+0.119+0.038)と同じである。また、回帰係数も単回帰の回帰係数と変わらない。  
280 これは、主成分が直交して、相関がないためである。ちなみに、単回帰の分散比の合計は、  
281 0.693 で、これが直交化した説明変数で、データの分散を説明した場合の説明できる分散  
282 の比であり、 $1-0.693=0.307$  が誤差分散の合計の比率である。主成分で重回帰すれば、  
283 個々の変数の説明力が、加算的に計算出来て、全体の説明力が上がる。しかし、数式の説  
284 明力を上げることは、言語的な説明力とは直接関係ない。PC1 は相関を排除した軸で、そ  
285 れは、A 豊かさ、B 都市化、C 識字率、D 工業化、E 政治的安定性にかかわる何かの要因  
286 であることは明らかだが、それが、Lipset(1960)が述べたところの近代化だとすれば、一  
287 見、Lipset(1960)の仮説が支持されたように見える。しかし、説明変数として採用したデ  
288 ータのすべてが同じ方向を向いていて、それと結果が正に相関しているので、近代化がも  
289 たらす何が、民主主義の維持に働いたのかわからない。この分析ではデータに用いなかった  
290 何かが民主主義の維持に働いているのかもしれない。問われるべきは、その何かが何か  
291 である。解説者は、浅学にして、かの高名な政治学者である Lipset の学説の詳細を知らない。  
292 彼の本にはそのメカニズムが詳述されているのかもしれない。もし、それが書かれて  
293 いるのであれば、その要因の程度を示す何らかのデータを加えて、それと、他のデータ項  
294 目、R の間で、偏相関分析すれば、豊かさ、識字率、政治の安定などの、近代化を媒介し  
295 ない部分と、R の相関を否定できるので、Lipset(1960)の仮説の検証になりえる。現時点  
296 では、第5主成分に明らかに豊かさにかかわる主成分があるので、第1主成分は豊かさだ  
297 けに関係する何かではないと考えるのみである。現時点では、Lipset(1960)の仮説を否定  
298 できないが、他のもっと根本的な原因がある可能性を否定することもできない。  
299

300 I I- 6. 因子分析 (R line124-147)

301

302 最後に、因子分析を行った。用いたデータセットは、表2のデータセットである。

303 因子分析では、最尤法で因子抽出を行い。Promax 回転で、軸の回転を行った。因子分  
 304 析は、因子数をデータ項目よりも少ない数に限定し、その因子数で、説明可能な分散を最  
 305 大化する分析である。目的は、因子数の限定であり、少ない数の因子に分散が集中し最大  
 306 化される。因子分析はそもそも、相関=0の制約を無視しているので、主成分分析のよう  
 307 に、直交した軸にはならない。さらに、promax 回転は、直交性を無視して、回転を行うの  
 308 で、因子間の相関が大きくなる。回転の目的は因子負荷をいくつかのデータ項目に集中さ  
 309 せて、因子の意味の解釈を容易にすることである。この操作によって、主成分分析の PC1  
 310 の内容が、いくつかに分離し因子の意味が解釈し易くなることを期待している。

311 表 8 に、因子分析の結果を示す。結果は期待通りに、第一主成分の内容を二つに分ける  
 312 形で、A:豊かさ、C:識字率、E:政治の安定と、B:都市化、D:工業化の2つのグループに分  
 313 かれた。つまり、工業化・都市化と近代化を分離したことになる。このことは図3のPC1-  
 314 PC2のプロットで、A、C、E、が上向きでまとまっており、B、Eが下向きでまとまってい  
 315 たことから、十分予測可能なことであったが、因子の promax 回転によって、「近代化」  
 316 が2つの因子に分かれたということは大きな成果である。もう少し細かく見ると、主成分  
 317 分析の PC1 ではA(豊かさ)の主成分負荷量が、C(識字率)やE(政治の安定)よりも高い値で  
 318 あったが、因子分析では因子1のAの負荷量をC、Eが上回っており、特にC(識字率)  
 319 の負荷量が、A(豊かさ)の負荷量を1.5倍ほど上回っていることも注目すべきである。  
 320 もしこの因子が、Lipset(1960)が言うところの、「近代化」であったとすれば、近代化は、  
 321 経済の発展よりも、より大きく一般庶民の識字率(教育の普及に)影響をもたらすと言え  
 322 るだろう。なお、結果の要約として、累積の分散比(0.712)が示されているが、この値には  
 323 全く意味がない。因子得点から計算した、FA1とFA2の相関係数は0.599であった。この  
 324 部分が重複している。2つ因子の分散の合計を説明力と受け取ることは誤りである。

325

326 表 8. 因子と R (民主主義の維持) の因子分析 (R line126-138) の結果の要約

	要約				負荷量	
	FA1	FA2			FA1	FA2
分散	1.858	1.701		A	0.665364	0.393248
分散比	0.372	0.34		B	0.040678	0.697954
累積	0.372	0.712		C	0.946746	-0.16137
				D	-0.02962	1.015819
				E	0.718941	0.038406

327

328 説明力を評価するために、回帰式を作って予測値と実際の値の間の誤差分散を計算し、説  
 329 明力を評価した。以下のその結果を示す。データセットは表9に示した。

330

331

単回帰

332

PR(>t)

分散比

333

R=0.5529FA1

0.0041\*\*

0.412

334

R=-0.01809FA2

0.931

0.00048

335

重回帰

336

R=0.8543FA1+0.4674FA2

0.617

337

PR(>t) 0.000187 \*\*\* 0.01253\*

338

単回帰分析の結果、FA1の係数は、0.553、危険率は0.0041で、この係数は統計的に有  
 339 意であり、全体の分散の41%を説明する。FA2の係数は、-0.0018で統計的には全く有意  
 340 ではない。また、Rの予測値はほとんど、実際のRの値とかけ離れており、全体の分散の  
 341 0.05%を説明するだけである。しかし、2つの因子で重回帰分析すると、FA1の係数は  
 342 0.854、FA2の係数は0.467となりともに統計的に有意な値となる。この係数の値は信用で

343

344

表9. 因子得点を使った回帰分析のデータセット

	FA1	FA2	R
AUT	0.132457	0.299984	-1.11631
BEL	0.30431	1.513244	1.240347
CZE	-0.75465	1.175544	0.868243
EST	1.204714	-1.97787	-0.74421
FIN	0.851779	-1.09541	0.496139
FRA	1.301073	-0.28362	1.240347
GER	-0.32362	1.170895	-1.11631
GRC	-1.70417	0.969633	-0.99228
HUN	-0.34578	-0.39953	-0.12403
IRL	2.045005	-2.43715	0.992278
ITA	-0.94714	0.636869	-1.11631
NLD	1.168683	0.190251	1.240347
POL	-0.09063	-1.41671	-0.74421
PRT	-2.38368	0.967647	-1.11631
ROU	-0.06602	-1.3536	-0.49614
ESP	-1.69413	0.752486	-0.99228
SWE	1.245663	-0.45887	1.240347
UK	0.056134	1.746197	1.240347

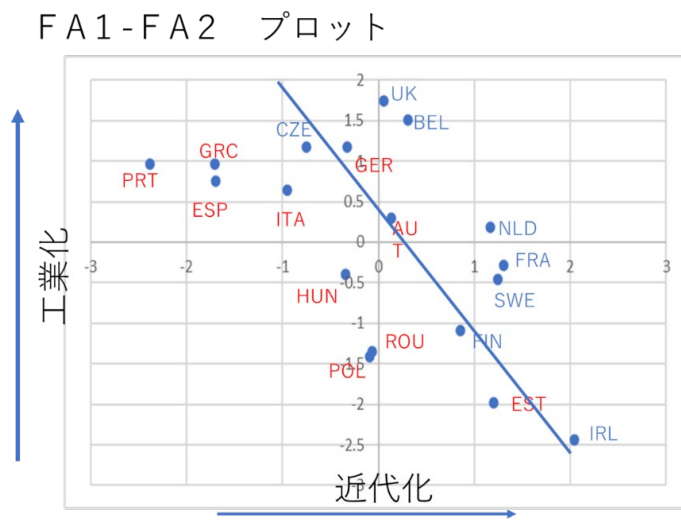
345



346 きない。2つの因子の間に相関があるため、多重共線性が疑われるからである。実際に式  
347 を使って、予測値を計算して求めた分散比は、0.617で、主成分分析から求めたPC1、PC2  
348 の重回帰の分散比、0.633より低く、説明力が低下している。因子分析の目的、特に  
349 promax回転の目的は、主成分分析で抽出され「近代化」の内容を、さらに分析的に分解す  
350 ることだったので、係数の推定値の信頼性を犠牲にして、近代化の内容を分解したこと  
351 なる。これは大きな成果である。しかし、この分析において、2つにわけた係数の信頼性  
352 は別の話である。

353 図5にFA1-FA2を軸とした散布図をしめした。図中に示した青線を境に、右上に民主  
354 主義維持国、左下に民主主義崩壊国が分布してる。主成分分析でPC1とPC2を散布図として  
355 プロットした図の左右を入れ替えた形で、境界線上にチェコスロバキア、フィンランド、  
356 アイルランド、ドイツ、オーストリア、エストニアが存在する。これはMDS(図2)、主成  
357 分分析(図3)の結果と同様得るが、この図の方が、境界線が鮮明である。この図の右  
358 上方向に、民主主義の維持の可能性が増大しているのである。主成分得点をプロットした  
359 分析よりも鮮明に結果を表している。それは大きな成果である。しかし、近代化は工業化  
360 の原因でまあり結果でもある。はたして、この議論において、近代化と工業化を分離する  
361 ことに意味があったのかという疑問は残る。

362



363

364

365

図 5.2 第一因子と第二因子の2次元プロット

366 II-7. 数値的解析結果の整理

367

368 数値解析で出来たことを整理しておく

369 1. 分析に使ったデータセットは、分析項目間の全てで、比較的高い相関があった。

370 2. MDS、主成分分析、因子分析、何れの分析でも、国の特性値の分布で、オーストリア、  
371 セコスロバキア、エストニア、フィンランド、ドイツ、アイルランドの6カ国が境界帯  
372 にあることが示された。

373 3. 主成分得点を使ったクラスタリングは、上記の6カ国が境界帯にあるという見方を支  
374 持した

375 4. 主成分分析では、全分散の67%を占める第1主成分と、17%を占める第2主成分が抽  
376 出された。第1主成分は、すべての分析項目とある程度の負荷量を持ちその内容は曖  
377 昧であったが、第2主成分は、教育レベル（識字率）と正の相関、都市化と工業化と  
378 負の相関を持ち、近代化した農業国を示すと思われた。

379 5. 因子分析・promax 回転によって、「近代化」を「近代化」と「工業化」に分離できた。

380 6. 標準化した元データ、主成分得点、因子得点を使って、結果（民主主義の維持）を目  
381 的変数とする、回帰分析を行った。その結果、元データでは、豊かさ、教育、政治の  
382 安定の係数が大きく、これら3つはすべて統計的に有意であったが、重回帰分析から  
383 予測され予測値と実データとの差の分散を全分散から差し引いて求めた予測値の分散  
384 （説明力）の合計は、100%を超えた。これは、データ項目間の相関が高いためであ  
385 る。この3つで重回帰すると、すべての係数で有意ではなくなった。項目間の相関が  
386 高く、データサイズが小さいために、このように結果になると思われた。その説明力  
387 は、全分散の61%を占めていた。

388 7. 主成分得点による回帰分析の結果、単回帰で、第一主成分の係数は有意で、その  
389 説明力は51%であった。第2主成分の係数は有意でなかったが、その説明力は12%で  
390 あった。また、第5主成分は分散が小さにもかかわらず、有意ではないものの第3、  
391 第4主成分よりも係数が大きく、説明力が4%あった。この3つの主成分を使った重  
392 回帰分析の結果は、第2主成分まで係数が有意で、式全体としての説明力は67%であ  
393 り、5つすべての説明力の合計は69%であり、元データの項目、因子得点を使った回  
394 帰分析よりも、説明力が高かった（理論的に当然）

395 8. 因子得点を使った回帰分析では、単回帰では第1因子の係数は有意で、41%の説明力  
396 があった、第2因子の係数は有意でなく、説明力もほとんどなかった。この二つを使  
397 った重回帰分析では、2つの係数がともに有意で、式の説明力も62%あり、第2因子  
398 が、第一因子を補完する形で、結果にかかわっていることが、推測された。

399

