

政治行政関連リレーショナルデータベース構築 のための基礎的研究

名取 良太^{*1} 鵜飼 康東^{*2} 岡本 哲和^{*3}

要 旨

本論文では、政治・行政データに関するリレーショナルデータベースの構築を目指した試みを紹介する。日本政治に関する多くの論文で統計分析が用いられる一方で、分析に用いるデータセットは、個々の研究者によってアドホックに作成されているのが現状である。この状況は、日本政治研究の発展を阻害するといつても過言ではなく、改善すべき点である。このような問題意識に立って、われわれは、さまざまな種類のデータを簡易な処理によって加工できるリレーショナルデータベースの構築を始めた。本論文が紹介するのは、その成果の一部である。

このデータベースの特色は、ドラッグ&ドロップによる簡易な操作性、Excelなど表計算ソフトによっても可能な入力形式の柔軟性、データ内容を確認しながら処理できる視認性を重視した点である。また、映像編集ソフトのタイムラインから着想を得たインターフェースを採用することも特徴の一つである。

着手を始めたばかりの取り組みであるが、本データベースの開発は日本政治研究の発展に貢献を果たすものと考えられる。

The Relational Database for Research on Japanese Politics

Ryota NATORI Yasuharu UKAI Tetsukazu OKAMOTO

Abstract

This paper presents our attempt to construct a relational database system designed for research on Japanese politics. Japanese political studies currently suffer from the insufficiency of data environment despite the growing uses of statistical analysis. As this can be a major factor hindering the development of research on Japanese politics, we began to construct a relational database system which utilizes an interface inspired by video editing software. The database system includes features that would facilitate data input and management in terms of convenience, visibility and flexibility. While there are still scopes for improvement, we believe that this system could contribute to the progress of the analysis of Japanese politics.

^{*1} 関西大学総合情報学部

^{*2} 関西大学総合情報学部

^{*3} 関西大学政策創造学部

1 イントロダクション

日本の政治学では、統計分析を用いた論文が急激に増加している。しかし、それとは対照的に、そのファンダメンタルズといえるデータ環境の整備は、いまだ疎かにされたままである。そこで我々は、このギャップを埋めるための作業、すなわち「政治行政関連リレーションナルデータベース」の構築を始めた。本論文は、その試みについて詳しく説明するものである。

政治学における統計分析の対象は、政治過程全般に及ぶ。選挙過程、政策立案過程、審議過程、政策決定過程、政策アウトプット、政策アウトカムとあらゆる場面が、その対象となる。分析対象となるアクターも、有権者、議員（候補者）、政党、官僚、そして利益集団まで、さまざまである。さらに分析のレベルも、国であったり、地方であったり、国際比較であったりする。また、アグリゲートデータによる分析もあれば、サーベイデータによるものもある。

ポイントは、分析対象こそこのように多様であるものの、利用するデータに共通性が高いことである。有権者の政治意識を分析する時には、サーベイデータだけでなく、選挙区の属性（たとえば投票率、候補者別得票率、候補者の個人的属性、選挙区を構成する市区町村レベルの社会経済環境）を示すアグリゲートデータを用いることがある。その選挙区の属性は、国会議員の議会における行動や、政党内部の組織構造や各議員の行動選択と関連付けられながら分析に利用される。補助金配分や公共投資の実態を分析する際には、（市区町村レベルの）社会経済環境がコントロール変数として用いられる。その社会経済環境は、地方政治の分析には欠かせないデータである。すなわち政治学では、各研究者が、さまざまな対象を“類似したデータを用いて”分析しているのである。そして、この点に、日本政治研究におけるデータ環境の問題点と、リレーションナルデータベース構築の意味が存在する。

その問題点とは何か。ここで、各研究者が、それぞれデータを収集・加工する状況——すなわち、それは現状であるが——を見てみる。もし各研究者の分析対象の相違が、収集するデータの相違をもたらすのであれば、それほどの問題は生じない。また、データの加工に、それほどの時間を必要としないのであれば問題は少ない。しかし上で述べたように、政治学においては分析対象が違っても、使用するデータが類似する事が多い。そして、後述するようにデータの加工には多大なコストを必要とする。また、そもそもデータ収集にも困難が多い。つまり現状は、個々の研究者が“類似データを、多大なコストをかけて収集・加工している”状態である。そして、この状態は、日本政治研究の発展を阻害する。

まず、研究に充てる時間を一定とするならば、収集・加工に時間を要するほど、分析時間が短くなる。ましてや全国各地で、類似したデータを同じように収集・加工しているわけだから、日本政治研究全体として失っている“延べ時間”は相当なものであろう。分析時間が長ければ良いというものではないが、何らかの知見を生み出す可能性を失うことは確かである⁽¹⁾。

つぎに、データの手入力が必要な場合には、エラーの発生する可能性が高い。エラーの多い

データによる分析では、結果の信憑（しんぴょう）性に問題がある。しかし、入力→エラーチェックという作業は、ますます分析時間を減少させる。さらに、このエラーチェックにかかるコストも、研究者それぞれにおいて発生していることも忘れてはならない。この問題も“延べ”で考えるべき点である。

加えて、データ収集の困難さは、①研究者の絶対数を減らし、②入手が容易なデータによる分析を増加させ、③“収集したこと”そのものに価値を付与させる。研究は、多数の研究者の競争によって発展していく。ところがデータの入手が困難であると、そもそも参入者が少なくなる。参入者は、分析自体を重視するので、不要なコストをかけたくない。そのため、容易に入手できるデータによって分析するようになる。しかし、時にそれは“分析目的と符合しないデータ”による分析をもたらせる。例えば“国庫支出金”に対する自民党議員の影響力を分析する時、都道府県レベルのデータ“のみ”を利用するようなことが生じる⁽²⁾。逆に、入手困難なデータを収集して分析する場合、競合者が現れず、理論的な競争が生じなくなる⁽³⁾。さらに個々の研究者がアドホックにデータを加工することは、上記データ収集の問題とも関連して、再現性のテストを困難にさせる⁽⁴⁾。このようにデータ環境の未整備は、政治学の発展の阻害要因となるのである。

これを解決するためには、日本政治研究の共有資源たるデータベースシステムの開発が不可欠である。そのシステムは、全国の研究者が同じ作業を繰り返すことを回避し、分析時間を著しく増加させる。エラーについても効率的な管理が可能になる。データ収集・加工が容易になれば、新規参入者の増加を促すし、他の研究者による再分析も可能になる。すなわち日本政治研究の発展に結びつくと考えられるのである。

以下では、リレーションナルデータベース構築の実際の試みについて、主にユーザー側の視点から説明していく。本論文は、データベースに関する技術的な論文ではない。あくまでも、日本政治研究者が、どのような種類のデータを、どのように加工したいと考え、どのような問題に直面しているのかを論じ、それを解決するために構築したデータベースの概要を提示するものである。

具体的な構成は、まず、従来のデータベース構築の試みについて簡単に述べ、そこでは何が足りなかったのかについて論じる。次に、データを加工する際に、いかに煩雑な作業が必要とされるかを説明する。次いでデータベース構築の基礎的作業といえる正規化が、本データベースではどのようになされたのかを提示する。その上で、実際のインターフェースが、どのように設計され、どのように作動するのかを説明していく。そして最後に、政治学におけるデータベースの課題と展望について論じることにしたい。

2 日本政治研究におけるデータ

日本政治研究に関して、本格的なデータ分析が始まられてから既に40年以上が経過してい

る⁽⁵⁾。その間、データ環境の整備がまったく疎かにされたわけではない。何人かの研究者は、収集したデータを、相互利用可能な形で提供している。

まずサーベイデータについてみると、大規模な有権者意識調査については個票データが公開されている。1976年のJABISS（三宅一郎・綿貫譲治・公平慎策・Flanagan・Richardson）、1983年のJES（猪口孝・三宅一郎・綿貫譲治・蒲島郁夫）、1993～1997年のJES2（三宅一郎・綿貫譲治・蒲島郁夫・小林良彰・池田謙一）、2002～2006年のJES3（池田謙一・小林良彰・平野浩）、1996年と2000年のJEDS1996及びJEDS2000（三宅一郎・田中愛治・池田謙一・西澤由隆・平野浩）、選挙ごとに実施される「明るい選挙推進委員会調査（明推協調査）」は、レヴァイアサン・データバンク（LDB）や東京大学社会科学研究所附属日本社会研究情報センター（SSJデータアーカイブ）、あるいは日本選挙学会のウェブサイトを通じて公開されている。

一方、アグリゲートデータで代表的なのは、水崎節文による総選挙データベース（JED-M）である。これは1958年に実施された第28回総選挙から現在に至るまでの、すべての総選挙について、市区町村別・候補者別にデータが入力されたものである。このデータは、現在でも、新たに総選挙が実施されるたびに更新される貴重なものである。その他、川人貞史・川人典子による「衆議院総選挙候補者選挙区統計1890-1990」、スティーブン・リードによる「衆議院議員総選挙全候補者得票結果」も公開されている。川人データは、第1回から第39回までの衆院選の選挙結果が候補者別に入力されるものである。候補者名・政党名・年齢・性別に加え、当選者得票数や最下位当選者得票数なども入力されており、さまざまな分析が可能なデータセットになっている。リードのデータセットは、時系列変化を捉えやすいことが特徴である⁽⁶⁾。さらに朝日新聞社や読売新聞社は、選挙が行なわれるごとに市区町村別・候補者別データを含んだCD-ROMを販売している。

このほか、国勢調査をはじめとする国が行う各種調査のデータは、Sinfonicaが提供・販売しており、丁・字単位でのデータ入手できる。地方財政のデータは、市町村レベルについては地方財務協会が発行する市町村別決算状況調（96年度版からCD-ROMが添付されるようになった）から、都道府県レベルについては総務省が発行する地方財政統計年報を用いて容易に収集できる。

このように、数人の研究者の努力、および政府からの提供によって日本政治研究を進めるためのデータは提供してきた。しかしながら、それらは各々の研究者や政府機関が独自のフォーマットで作成したものであり、相互の関係性が意識されていない。そのため、たとえば意識調査データと候補者情報、及び国勢調査データを結合した分析を行なうためには、相当なコストを支払わねばならない。

それでは、異なる種類のデータの結合を意識したシステムが存在しないかといえば、そうではない。日本経済新聞社が提供する日経フィナンシャルエスト（旧日経NEEDS）は、国勢調査や地方財政をはじめとする各種データを収納し、ユーザーが自由にデータ項目を選択し、ダウンロードできるデータベースである。しかし、このシステムは社会・経済データのみを扱

っており、選挙データや行政データを含んでいない。そのため、選挙データを扱った場合に発生する特殊な問題に対応するためには、やはり相当な作業コストが必要である。

また、未公開ではあるが、小林良彰による慶應義塾大学政治データバンクシステムは、政治意識調査データと選挙データに加え、社会経済データや財政データまでを収録した大規模なシステムである。このシステムでは、データの正規化が施されると共に、インターフェースにも工夫が加えられている。その結果、ユーザーは簡便に、指定したデータを任意の形式で取り出せるようになっている。ただし、このシステムでは異なる種類のデータを結合することができない。

以上見てきたように、日本政治研究を進めるためのデータ環境は、データそのものこそ存在しているものの、各々の研究者が多くのコストを支払わねば分析が進められない状態である。もちろん、データそのものが存在しなければ分析すら不可能であるから、先達の貢献には敬意を表さねばならない。しかしIT技術の飛躍的発展を見せる現在、これらの業績を一方的に受け入れつづけるのではなく、より発展させることが求められよう。したがって、われわれは、これらのデータを統合し、加工可能なデータベースシステムの開発を試みるのである。

3 政治関連データの加工にかかるコスト

さまざまなデータを結合することによって分析用のデータセットを作成する時、その作業は非常に煩雑になる。そのイメージをつかんでもらうため、名取^[1]で行われた分析に要した作業を例に挙げて説明したい。

この論文は、選挙制度改革が補助金配分に与えた影響を分析している。そこでは、自民党の強い地域ほど多くの補助金が配分されるという特徴が、小選挙区比例代表並立制の導入によって変化したのかどうかを検証している。では、これを検証するためのデータセット作りに、いかなる作業が必要だったのか。

はじめに、市区町村別・選挙区別の自民党得票率などを算出する⁽⁷⁾。これにより、その地域における自民党の強さが、数値で表現できるからである。そこで前述のJED-Mデータを用いて、これを算出した。

レヴァイアサン・データバンクから購入したJED-Mデータには、選挙区ごとに分割された選挙結果データがCSVないしテキスト形式で収められている。具体的には、中選挙区制で実施された第28回から第40回のデータには129ないし130のファイルが、小選挙区比例代表並立制で実施された第41回以降のデータには300小選挙区選挙結果と、300小選挙区別の比例代表選挙結果ファイルが収められている。

では、これらのデータの中から一つのファイルを選択してみよう。なおここでは、プログラム作成などについて特別なトレーニングを受けていない、文系学部出身の研究者や文系学部に所属する学生でも利用できるMS Excelによって処理する場合を考えていく。図1は第39回選

挙の北海道一区のデータである。JED-Mデータの利点は、すべてのファイルが、同じフォーマットで入力されていることである⁽⁸⁾。北海道1区でも、大阪3区でも、C列6行目には当該選挙区の有権者総数が入力されている。そのため、簡易なプログラムを組んでしまえば、すべての選挙区について簡単に処理できる⁽⁹⁾。ここでの処理は、市区町村別の自民党得票率を算出することである。したがって、各ファイルの自民党候補者の得票データを取り出し、投票者数で除するようなプログラムを書けば良い。ただし、中選挙区制の下では、一選挙区に複数の自民党候補者がいるので、得票数の“合算”をプログラムに組み込む必要がある。ともあれ、入力フォーマットが統一されているおかげで、市区町村別自民党得票率を算出するのは比較的容易である。なお、簡単なプログラムが組めない場合、途方もない手作業が必要になる。

話は少し横道にそれるが、朝日新聞社が販売する『asahi.comで見る2000年総選挙のすべてCD-ROM版』のデータを使用すると、一段階多くの行程を踏まねばならない。asahi.comの小選挙区データは、一つの都道府県内の全選挙区のデータを、一つのファイルにまとめている。したがってファイル数は、北海道から沖縄までの47ファイルとなる。図2は、北海道のデータファイルの一部である。図からわかるように、このファイルでは、選挙区の変わり目が不規則である。何行目に北海道1区から2区に移り、2区から3区に移るのかに規則性が無いのである。これは、選挙区が変わるタイミングが、各選挙区が有する市区町村数に依存しているためである。したがって、プログラムには、次の選挙区データになったことを識別するような仕掛けを組み込まねばならない。この分だけ、JED-Mに比べて余分な行程が必要になる。

さて、いずれのデータを用いても、簡単なプログラミングの知識さえ有していれば、市区町村別の自民党得票率を算出するのは容易である。その理由は、前節でも述べたように、ここで扱っているデータが「一主体によって作成された、共通性の高いファイル内部での加工作業」だからである。われわれがデータベース作成の必要を強く感じるのは、この先の作業を始めてからである。

つぎに必要な作業は、補助金の配分状況を示すデータを算出することである。具体的には、市区町村別の財政データを用いて、歳入に占める国庫支出金収入の割合を算出したり、その割合を他の市区町村と対比したり、前年度と対比するなどして指標を作成することである。市区町村レベルの財政データは、地方財務協会が発行する「市区町村別決算状況調」および日経FQから入手できる。この二つのデータの違いは、項目数と電子データ化されている年度数である。「市区町村別決算状況調」には250近い項目が掲載される（図3）。国庫支出金の内訳まで細かくカバーしているので、政治家の影響が及びにくい国庫負担金額を差し引いた指標を作成できる。しかし、「決算状況調」の電子データ化は、1996年度以降に限定される。そのため中選挙区制度の影響を分析するには不十分である。中選挙区制での選挙は、1993年までしか実施されていないからである。これを分析するためには、膨大なデータを手入力しなければならない。

これに対して日経FQは、1975年度分データから電子化されており、手入力なしで選挙制度改革に対応する事ができる（図4）。ただし、入力されている項目数が、「決算状況調べ」に比

選挙区名	地図数	有権者数	投票者数	有効投票数	候補者数	有権者数	投票者数	有効投票数
40:北海道1区	30	6	12	1905093	1136672	1133949		
41:								
42:								
43:								
44:								
45:								
46:								
47:								
48:								
49:								
50:								
51:								
52:								
53:								
54:								
55:								
56:								
57:								
58:								
59:								
60:								
61:								
62:								
63:								
64:								
65:								
66:								
67:								
68:								
69:								
70:								
71:								
72:								
73:								
74:								
75:								
76:								
77:								
78:								
79:								
80:								
81:								
82:								
83:								
84:								
85:								
86:								
87:								
88:								
89:								
90:								
91:								
92:								
93:								
94:								
95:								
96:								
97:								
98:								
99:								
100:								
101:								
102:								
103:								
104:								
105:								
106:								
107:								
108:								
109:								
110:								
111:								
112:								
113:								
114:								
115:								
116:								
117:								
118:								
119:								
120:								
121:								
122:								
123:								
124:								
125:								
126:								
127:								
128:								
129:								
130:								
131:								
132:								
133:								
134:								
135:								
136:								
137:								
138:								
139:								
140:								
141:								
142:								
143:								
144:								
145:								
146:								
147:								
148:								
149:								
150:								
151:								
152:								
153:								
154:								
155:								
156:								
157:								
158:								
159:								
160:								
161:								
162:								
163:								
164:								
165:								
166:								
167:								
168:								
169:								
170:								
171:								
172:								
173:								
174:								
175:								
176:								
177:								
178:								
179:								
180:								
181:								
182:								
183:								
184:								
185:								
186:								
187:								
188:								
189:								
190:								
191:								
192:								
193:								
194:								
195:								
196:								
197:								
198:								
199:								
200:								
201:								
202:								
203:								
204:								
205:								
206:								
207:								
208:								
209:								
210:								
211:								
212:								
213:								
214:								
215:								
216:								
217:								
218:								
219:								
220:								
221:								
222:								
223:								
224:								
225:								
226:								
227:								
228:								
229:								
230:								
231:								
232:								
233:								
234:								
235:								
236:								
237:								
238:								
239:								
240:								
241:								
242:								
243:								
244:								
245:								
246:								
247:								
248:								
249:								
250:								
251:								
252:								
253:								
254:								
255:								
256:								
257:								
258:								
259:								
260:								
261:								
262:								
263:								
264:								
265:								
266:								
267:								
268:								
269:								
270:								
271:								
272:								
273:								
274:								
275:								
276:								
277:								
278:								
279:								
280:								
281:								
282:								
283:								

①都市別 人口内訳						
都道府県	市町村名	地方内訳				
		地方税	市町村民税個人分	市町村民税法人分	固定資産税	市町村たばこ税
北海道	札幌市	256,836,442	79,947,896	29,642,985	111,198,933	14,992,886
北海道	北広島市	32,142,153	10,978,459	3,141,744	3,318,372	2,373,375
北海道	小樽市	16,079,729	4,581,427	1,313,960	7,408,385	1,135,341
北海道	旭川市	40,623,994	12,153,728	4,132,236	6,471,368	2,991,638
北海道	室蘭市	14,651,859	3,269,827	1,336,995	7,897,254	846,598
北海道	函館市	23,173,487	6,612,751	2,363,463	10,255,971	1,673,610
北海道	恵庭市	21,614,995	7,011,585	2,167,402	8,866,619	1,595,117
北海道	北広島市	12,424,397	3,764,585	1,623,394	5,018,930	936,795
北海道	夕張市	1,067,067	31,988	72,748	519,534	106,853
北海道	増毛沢市	7,652,546	2,515,209	717,025	3,315,163	656,709
北海道	網走市	4,846,472	1,706,153	475,600	1,938,900	392,992
北海道	勇払市	3,034,394	1,039,692	302,545	1,192,975	262,054
北海道	空知郡新得町	29,146,716	5,757,171	3,000,566	5,806,449	1,530,619
北海道	勇払郡内浦町	4,769,913	1,733,387	485,946	1,742,227	452,164
北海道	勇払郡美幌町	2,354,135	766,461	161,077	1,021,913	237,730
北海道	勇払郡勇払町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡豊頃町	11,623,245	4,033,056	528,609	5,074,799	729,813
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,980,726	911,145	309,188	1,167,026	269,784
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡勇払町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡豊頃町	1,056,442	253,672	85,332	609,141	88,136
北海道	根室市	3,039,049	1,181,206	303,555	1,024,982	333,918
北海道	千歳市	13,715,580	3536,495	1,223,030	6,947,267	799,338
北海道	釧路市	4,649,300	1,446,773	441,305	1,937,192	412,905
北海道	稚内市	2,269,489	621,248	216,240	1,108,603	181,137
北海道	恵庭市	256,006	101,989	29,191	80,473	35,201
北海道	勇払郡富良野町	2,421,475	762,415	193,898	1,071,727	228,072
北海道	勇払郡富良野町	2,296,836	702,525	209,518	1,012,776	199,874
北海道	勇払郡美幌町	5,317,957	1,509,483	285,945	2,353,264	422,469
北海道	勇払郡上士幌町	2,269,489	621,248	216,240	1,108,603	181,137
北海道	勇払郡豊頃町	1,056,442	253,672	85,332	609,141	88,136
北海道	勇払郡中頓別町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡日高町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡上士幌町	2,980,726	911,145	309,188	1,167,026	269,784
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583
北海道	勇払郡上士幌町	2,155,732	592,126	234,036	995,996	189,840
北海道	勇払郡豊頃町	1,810,819	401,513	89,706	1,051,387	154,903
北海道	勇払郡中頓別町	1,025,733	322,583	88,945	432,451	123,202
北海道	勇払郡日高町	2,685,653	963,546	277,274	1,030,238	217,583</td

べて少ない。そのため分析対象によっては、手入力によってデータを補う必要がある。

とはいっても、指標の作成にまったく面倒はかからない。いずれのデータセットにおいても、一つの行に、一つの市区町村の全財政データが入力されているからである⁽¹⁰⁾。したがって、単純な割合や全国平均比の算出は、なんら工夫のない数式を入力するだけでよい。これも選挙データと同様、「一つの主体によって作成された、一つのフォーマットによるデータ」だからである。

データ加工の問題は、ここからである。補助金と自民党得票率の関係を分析するためには、市区町村別の選挙データと補助金データを結合しなければならない。たとえば札幌市の補助金データを、札幌市の選挙データ結合しなければならないのである。

複数のデータセットを一つのデータセットへと結合するためには、結合の基準値が必要である。共通する基準値があれば、ある基準値を持つ補助金データを、選挙データのうちの同じ基準値を持つ行に結合すればよい。ところが、この基準値に大きな問題がある。ここでもう一度、図1～図4までを見てみよう。

この4つの異なるデータセットに共通する項目は「自治体名」である。しかし、自治体名が同じである行を、「同じ自治体」として認識させてはいけない。同じ名前の別の自治体が存在するからである。たとえば中央区は札幌市、東京都、大阪市、神戸市にある。さらに言うと、群馬県の東村や兵庫県の一宮町のように、同一県内に同じ名前の自治体があるケースもある。したがって、"文字列"を用いて結合するには、「都道府県名」「市郡名」「区町村名」を組み合わせねばならない。このため、全てのデータファイルに、あらためて都道府県名と郡名を加えるという作業が必要になる。

ただし、この方法は、推奨できない。第一の理由は、エラーが発生しやすく、かつチェックが困難なことである。第二の理由は、その後のデータ処理に適さないことである。

そこで利用するのが、JISコードである。総務省統計局は、1970年に地方自治体に対してJISコードを付与した。これにより都道府県名・市郡名・区町村名を合わせてひとつの数値で表現できるようになった。したがってJISコードと自治体名の一覧表を入手すれば、文字列である「自治体名」から数値である「JISコード」を抽出し、それを基準として結合できるようになる。もちろん、この抽出作業においても、先述の「同一自治体名」の問題を避けることはできない。違う都市の"中央区"のコードを抽出してしまう恐れがある。そのため、自治体名からJISコードを正しく抽出するには、やはり都道府県と郡という二つのレベルでフィルターをかけねばならない。そのため手間という点から見れば、文字列を使って結合するのと同じである。ただし上で指摘した二つの問題、エラーチェックと後のデータ処理を考慮すると最終的な手間は少なくて済む⁽¹¹⁾。

さて、この作業により、全てのデータテーブルにJISコードが入力された。そのため、あとはJISコードを基準にしてデータを結合させるだけで、目的であるデータテーブルが完成するはずである。ところが、このままでは多くの市区町村についてNULLが発生する。実は、選

選挙データに存在する自治体が、国勢調査や地方財政データには存在しないケースがある。たとえば札幌市や仙台市などの政令指定都市に設置される行政区は、独自の予算を作成しない。そのため地方財政データの中に、行政区は存在しない。また小選挙区制の下では、千葉県松戸市・大阪府堺市のように、市区町村が二つに分割され、異なる選挙区に属させるケースもある⁽¹²⁾。こうした選挙データの特質や、行政区の特質（選挙や一定の社会経済データについては測定単位になるが、財政については測定単位にならない）を原因とする問題には、個別的な対処を余儀なくされる。この論文では財政データ（国庫支出金）を従属変数にしているため、選挙データの和を、もとの市区町村のデータとみなすことにした。もし選挙データを従属変数にする場合には、財政データを行政区ごとあるいは分割された市区町村ごとの人口によって按分し、行政区・分割区のデータとみなすことになる⁽¹³⁾。そして重要な問題は、これらの作業は手作業で埋めていかざるをえない点である。

ともあれ、こうした作業を通じて、分析のためのデータセットが完成する。ここから、ようやく本当の作業である分析に取り掛かれるのである。とは言うものの、「この程度の手間は、大きな手間とはいえない」と見る向きもある。しかし、各研究者がそれぞれバラバラにこのような機械的作業に煩わされるのは損失である。

また、データセット作成の作業は、対象年度を増やしたり、対象となる選挙を広げたりすると膨大なものになる。対象年度を増やしたときに、まず生じる問題は、JISコードの問題である。地方公共団体コードは1970年にJISに指定されたので、1970年までに消滅した市区町村にはJISコードが与えられていない。すなわち1970年以前の選挙や財政を分析する時には、何らかの形で市区町村にコードを与えなければ、データセットを統合できない⁽¹⁴⁾。また70年代以降のデータを扱う場合でも、多くの“特異な事態”に対処せねばならない。たとえば、大阪市北区は、1989年に北区と大淀区を吸収合併した。現在では、こうした吸収合併の場合、名前を変更しない自治体についてはJISコードも変更されない。しかし大阪市北区は、新しい北区として、新しいJISコードが与えられた。この“同じ都道府県名、市郡名、自治体名であるのに異なるJISコードが付与される”という問題には、どうしても個別的な対処を余儀なくされる。

また市町村合併については、区の分割や新設というケースも考慮に入れねばならない。最近の例では札幌市の厚別区、横浜市の都築区がそれにあたる。この場合、通常の市町村合併では旧行政区域のデータを合算すればよいが、それを適用することができない。先の、小選挙区における分割区と同じように、過去のデータを、現在の行政区域の人口比で按分するという作業が必要である。

都道府県議会選挙を分析する際には、自らの手でデータの収集からはじめねばならない。データ収集と入力作業には、森^[2]にも示されるように、周到な準備と根気が必要である。各都道府県選挙管理委員会への依頼から始め、データクリーニングを終えるまでに、相当な日数を必要とする。収集すると明らかになるのは、都道府県ごとにデータ入力のフォーマットがバラバラなことである。また、分析ではなく、公開を目的としたデータ作りを行なっているので、

その後の処理ではなく見栄えを重視したフォーマットになっていることが多い（たとえばExcel上でセルの結合などが行なわれる）。また制度上の特別な現象として、選挙時には合併済みであるにもかかわらず（年月日を起点にしたコーディングでは合併しているにもかかわらず）、実際の選挙は、合併前の行政区域で実施されているケースがある。このようなケースには、やはり個別的な対応が必要とされる。

このほか、社会経済データを扱うときにも気をつけねばならない点がある。日経FQのデータは、実に調査に忠実である。それゆえに、市町村合併が生じた自治体について、4月に行なれた調査に対しては旧自治体ごとにデータが付与されているが、10月に実施された調査については新自治体にデータが付与されることになる。一方で財政データは、つねに当該年度末を基準に発表される。したがって、市町村合併が生じたとき、一部（合併前の当該年度中に実施された調査）のデータが欠落することになる。

以上見てきたように、社会経済データと選挙データを結合させてデータセットを作成するときには、さまざまな問題に直面することになる。そして、これらの問題は、こうしたデータを用いて分析しようとする全ての研究者に対して、共通に発生するものであり、それぞれが解決するためにコストを支払うことになる。

なお、ここまでアグリゲートデータ間の結合する際の問題を中心に論じてきたが、この問題は、サーベイデータにアグリゲートデータを結合する時にも同じように生じる。さらに、サーベイデータについては、独自の問題が発生していることも、最後に指摘したい。それは回答の選択肢に対するコードの振り方の問題である。たとえばJES2は、数年間にわたり、同じ有権者に繰り返し調査を行なうパネル調査を行なっている。そして、この調査が行なわれたのは、激しい政界再編が行なわれた時期であり、政党の離合集散・設立・消滅が頻繁に生じた時期である。ところが蒲島・綿貫・三宅・小林・池田^[3]によると、各政党に対して共通のコードが振られていないことがわかる。表1は、第一波と第七波における、ある設問に対する選択肢の

表1

JES2 第一波	JES2 第七波
1 自民党	1 自民党
2 社会党	2 新進党
3 公明党	3 民主党
4 民社党	4 社民党
5 共産党	5 共産党
6 社民連	6 新党さきがけ
7 新生党	7 自由連合
8 新党さきがけ	8 新社会党
9 日本新党	9 民改連
10 その他	11 無所属
11 無所属	
99 名前あげず	0 非該当

*蒲島郁夫・綿貫譲治・三宅一郎・小林良彰・池田謙一『JES II コードブック（変動する日本人の選挙行動⑥）』より筆者作成。

コードである。一見してわかるように、そこに共通性はない。第一波での「2」が社会党であるのに対し、第七波での「2」は新進党である。そのほかのコードについても、各回で対応する政党が異なっている。すなわち、政党名を識別する共通のコードを使用せず、アドホックに番号を与えているのである。このような処置が、後の分析において、一つの障害になることは容易に想像がつくだろう。

それでは、これらの問題を解決するために、我々がどのような点を考慮し、データベース構築作業を進めていったのかについて、節をあらためて論じていくことにしたい。

4 正規化とインターフェース構築

リレーションナルデータベースとは、複数のテーブル間のリレーションを構築し、データを呼び出すシステムである。このシステムを構築するに際し、もっとも初歩的な作業が正規化である。正規化を進めることにより、データテーブル間に重複した情報が収まることを回避し、効率的な検索が可能になるからである。そのためリレーションナルデータベースに関する教科書やマニュアルでは、正規化について丹念に記述されることがしばしばである。

本データベース構築に当たっても、もちろん、重複する情報がテーブル内に入力される事は回避しなければならない。しかしながら、政治行政関連データを扱う際には、別の観点からの正規化作業が必要とされる。政治行政関連データでは、一つのデータに対して非常に多くの情報が詰め込まれているので、その情報を適切に分別していかねばならない。隠されている重要な情報をあぶりださねばならないのである。不要な情報をそぎ落とすのではなく、必要な情報を明示し、むしろ付け加えていかねばならないのである。そこで具体的な例を挙げて、この点を確認してみよう。

表2は、有権者数と候補者別得票数に関するデータファイルを作成するために行なった、正規化作業の結果である。各項目は、ある数値が、どのような意味を持っているかを識別するために必要な項目を示している。ある数値は、どの選挙単位（衆院・参院・知事・都道府県議会・市区町村長・市区町村議会）の、どの選挙タイプ（選挙区・全国区・比例区か）の、どの投票

表2

有権者数	得票数
electionCode	electionCode
electionType	electionType
votingType	votingType
times	times
prefectureCode	prefectureCode
cityCode	cityCode
electorate	candNumber
	numberOfVote

タイプ（拘束名簿式・非拘束名簿式・単記式・連記式）の、第何回の、どの県の、どの市区町村（投票行政単位の）ものであるか、が決定されることによって、意味が付与される。得票数についてのある数値は、それに加えて、何番目の候補者であるかの情報を加えることによって、初めて意味が特定されるということである。

この正規化作業を進めるにあたって最も考慮した点は、とくに選挙関係データについては予測不能な事態が生じる可能性があるということである。たとえば参議院選挙では、1983年選挙からそれまでの全国区に変わって比例代表制が導入され、2001年選挙からは同じ比例代表制であっても拘束名簿式から、非拘束名簿式へと制度が変更された。このような制度変更は、データの識別作業に、多くの不備を発生させることになる。したがって、制度変更が生じた際に、その変更に伴う要素を付け加えるだけで（それまでの付与していた項目は一切変更することなく）対応できるように注意を払った。

ここで挙げたのはほんの一例であり、実際には、選挙データの正規化を進めた結果、25個のデータファイルを必要とした。また、社会経済データや調査データの正規化作業においても、それぞれ、多くのデータファイルを作成した。

この作業の一方で、インターフェースの仕様についても検討を進めた。まず基本とするのは、ユーザーの特性を考えれば、簡易な操作性を備えることが望ましいという点である。したがって、ドラッグ＆ドロップ、あるいはクリックのみでデータが抽出できることを目指した。

つぎに異なる種類のデータファイルを結合する点については、その可能性と制約条件を中心に考えを進めた。可能性については、データは縦方向にも横方向にも結合するというものである。前節で論じていた結合上の問題は、ある時点の選挙データと社会経済データを横方向につなぐ際に生じるものである。しかし、分析対象によっては、異なる時点のデータを用いることがある。たとえば、堀内・名取^[4]は、衆院選の選挙結果に対する都道府県議選の影響を分析した。この分析に使用するデータセットは、1996年、2000年、2003年衆院選をベースにし、それぞれに直近の都道府県議選のデータを結合させたものである。すなわち、もともとバラバラに存在していた3回の衆院選データを縦につなぎ、それぞれ横に都道府県議選データをつなぎデータセットである。したがって、こうしたデータセットを作成する需要があることを考慮する必要があった。

一方、制約条件については、なるべくそれを排除するように心がけたが、一つだけ設定せざるを得なかった。それは、データ結合の基準をユーザーが設定することである。換言すると、ユーザーは、データセットを抽出する前に、従属変数を確定しなければならないということである。衆院選の結果を分析するのか、参院選なのか、はたまた都道府県議選なのか、何年に実施された選挙なのか、単位は市区町村なのか、選挙区なのか、あるいは財政データを従属変数にするのか、これを特定しなければならない。

前節で述べたように、選挙を従属変数にするのか、財政を従属変数にするのかで、データ入力の単位は異なる。選挙データには、分割選挙区が存在するためである。したがって、選挙を

従属変数にした場合は、財政や社会経済データを按分しなければならないし、逆に財政を従属変数にした場合は選挙データを合算しなければならない。また、何年のデータを従属変数にするかによっても、抽出すべきデータは異なる。現在を基点にする場合、合併した自治体の過去のデータは、構成自治体の過去のデータを合算すれば対応可能である。しかし、過去を基点にして現在のデータを用いた分析をしようとするとき、合併自治体をサンプルからドロップするか、もしくは按分を施す必要がある。こうしたことから、「分析の基準」を特定するという制約だけは、ユーザーに課さざるを得ないという結論に至った。

さて、これらの点に加えて、データ入力の柔軟性という点にも目を向けた。本節のはじめに、データの正規化作業について説明した。正規化作業はリレーションナルデータベースを構築する際の基本的な作業ともいえる。しかしながら、元となるデータを作成する主体（たとえば選挙管理委員会）に、正規化されたデータの作成を要請することは不可能である。また、個々の研究者が、独自にデータを収集・入力するときは、時間効率を考えれば、一旦正規化するよりも、直接分析可能なデータセットを作成する方が望ましい。そのため、個々の研究者に対して、その作業を要請することは困難である。かといって、それらのデータを入手した後、あらためて正規化を行なうことも非効率である。そこで、十分に正規化されていないデータファイルであっても、データベースに取り込むことができるかどうかについて、その可能性を探り、それを実現できるようなシステムの構築を目指した。

以上まとめると、データベースの構築にあたっては、その基本となる正規化作業を注意深く進めつつ、操作の簡便性、データ抽出の自由度、データ入力の柔軟性を高めるようなシステムの構築を目指したのである。次節では、現時点で完成しているシステムについて、実際の処理手順を示しながら紹介していくことにする。

5 データベースシステムの紹介

処理手順の説明に移る前に、インターフェースの設計に際し参考にしたものを見たい。本データベースシステムの重要な特徴は、異なる種類のデータファイルを縦方向と横方向に結合するという点である。したがって、インターフェースを設計する際には、これと同じような特徴を持つ何かを参考にすべきと考えた。そこで目をつけたのが、映像編集ソフトである。映像編集ソフトは、いくつかの画像（動画）を連続的（横方向）に結合させ、さらに種々のエフェクトを被せていく（縦方向に結合させる）ことができる。この特徴は、まさに、われわれが構築しようとするシステムと適合する。また、ほとんどの操作をドラッグ＆ドロップで行なえる簡便性も持ち合わせている。このような理由から、インターフェース設計の基本的な部分は、映像編集ソフトから着想を得て構築することになった。なお、このことが、思わぬ副産物をもたらせたことも付記しておきたい。それは抽出データに関する視認性を、システムに付け加えることができたことである。すなわち、選択したデータの内容を確認しながら操作することを

可能にさせたのである。

それでは、本データベースの画面を示しながら、実際の処理手順について説明していくことにしたい。図5は、起動後の画面である。左側上部にはデータファイルの一覧が示される。本論文中に示している画面はテスト用のため、有権者数および市区町村コードのファイルのみを掲載しているが、実際には、ここにさまざまな種類のファイルが格納・表示されることになる。

つぎに、必要とするデータファイルを右側上部の枠内にドロップする（図6）。今回の試作

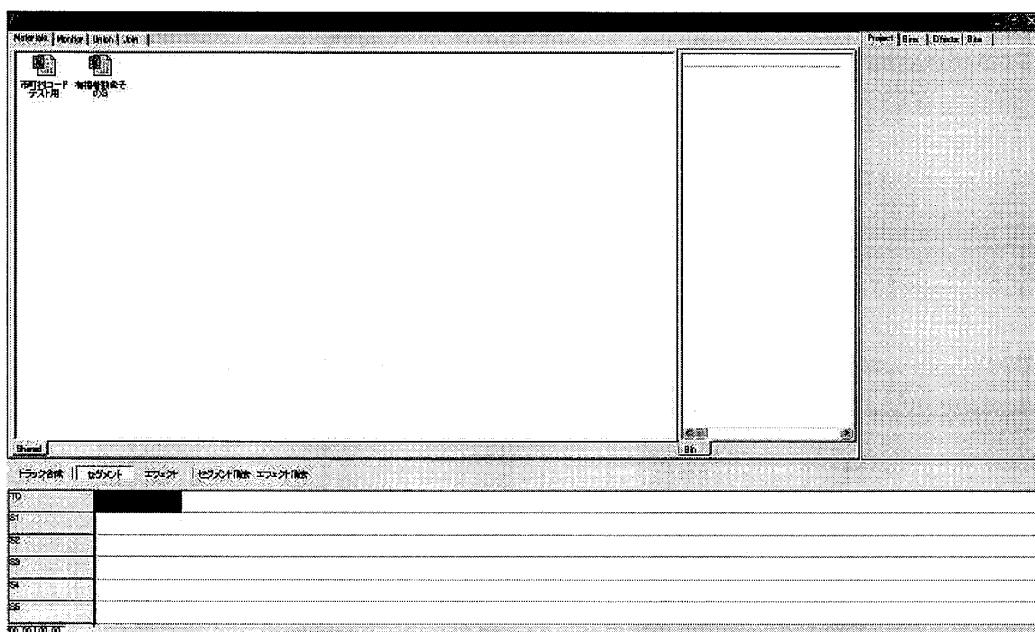


図5

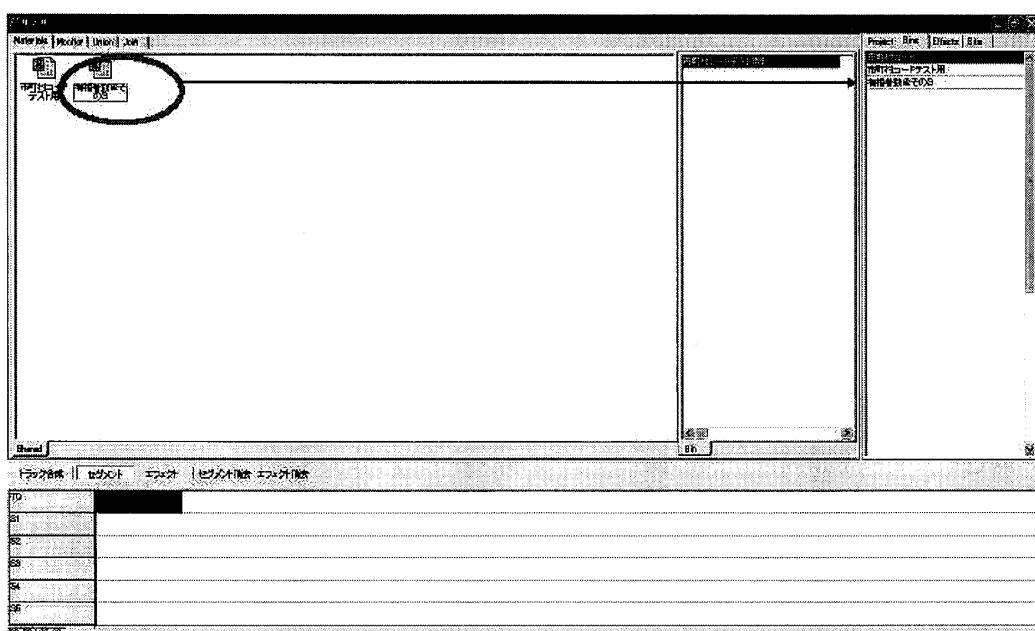


図6

では、Excelファイルに対応するように構築したので、Workbook内の各Sheetからデータを取得することが可能になる。右側上部の枠内でデータファイルをダブルクリックすると、中央上部の枠内にSheet一覧が表示される（図7）。同時に、左側上部には当該Sheetに入力されているデータの一部も表示され、中央上部の枠内には入力されているデータの全項目も表示される（図8）。

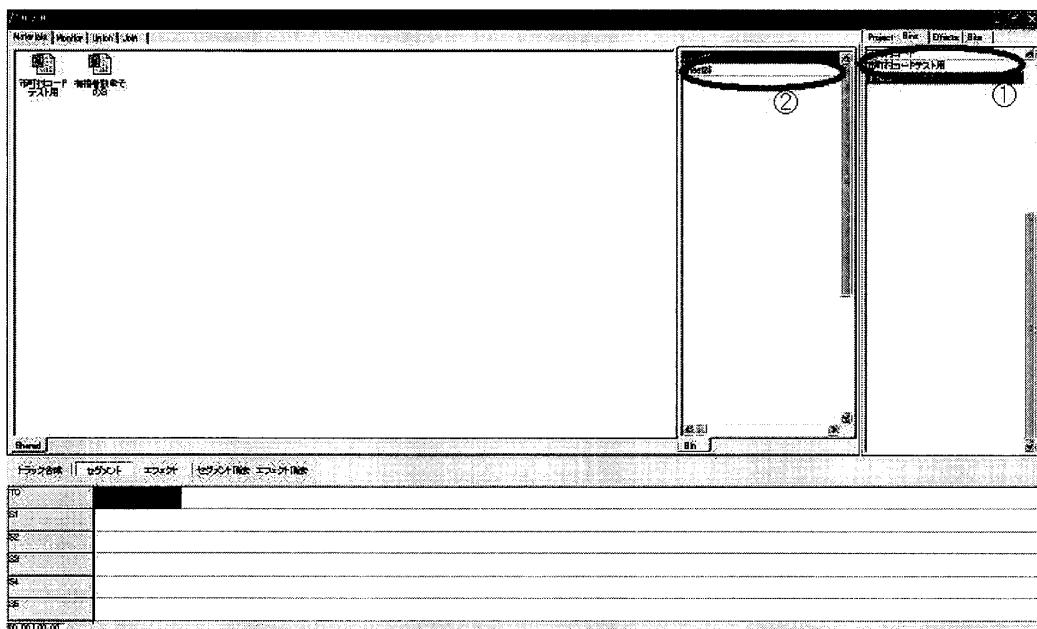


図7

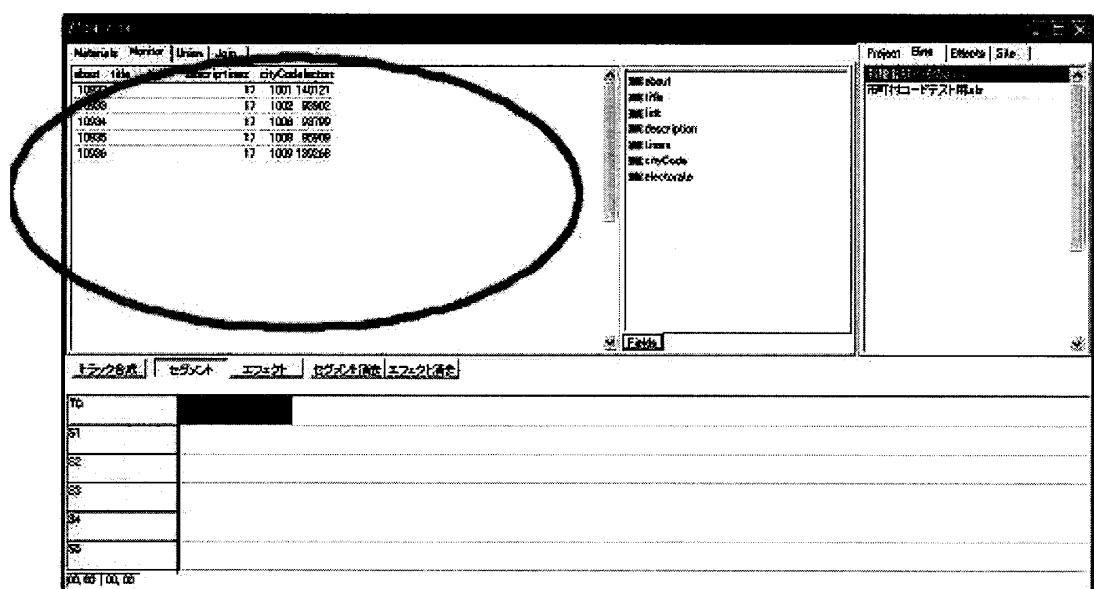


図8

このような手続きで取得データを選択したら、左側上部のデータ表示画面からタイムラインにドロップする（図9）。ここでは1993年衆院選のデータをベースにし、1995年参院選のデータを、市区町村単位で結合して取得するという例を示す。Sheet 1には93年衆院データ、Sheet 2には95年参院データが入力されているので、それぞれトラック1とトラック2にドロップする（図10）。結合する単位は市区町村なので、市区町村コードをトラック3にドロップする（図11）。

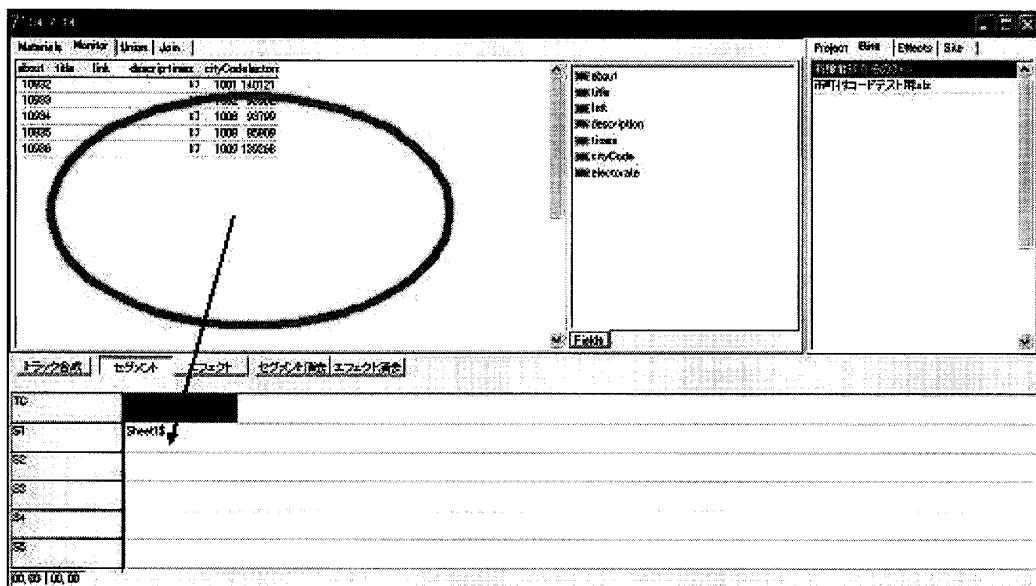


図9

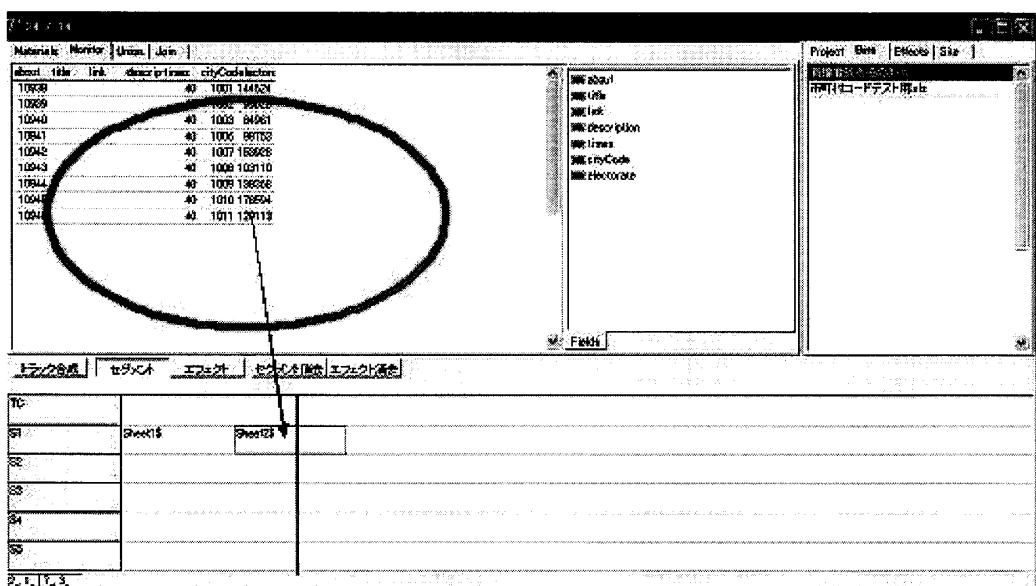


図10

一方、前節で示した、異なる時点のデータを用いた分析を行なう場合には、縦方向にデータファイルをドロップすれば良い。前節の堀内・名取^[4]で用いられたデータセットを作成するためには、96年衆院選データの下に、2000年衆院選データが入力されたシートをドロップ、さらにその下に2003年データのシートをドロップし、それぞれに対応する都道府県議データを横にドロップすれば作成できる。

さてタイムライン上に必要なデータをドロップした後は、結合のための処理をしなければならない。処理手順については、また改善の余地が残されていることを断らねばならないが、基本的な構造について示していくことにしたい。まず、トラック合成ボタンをクリックする（図11）。

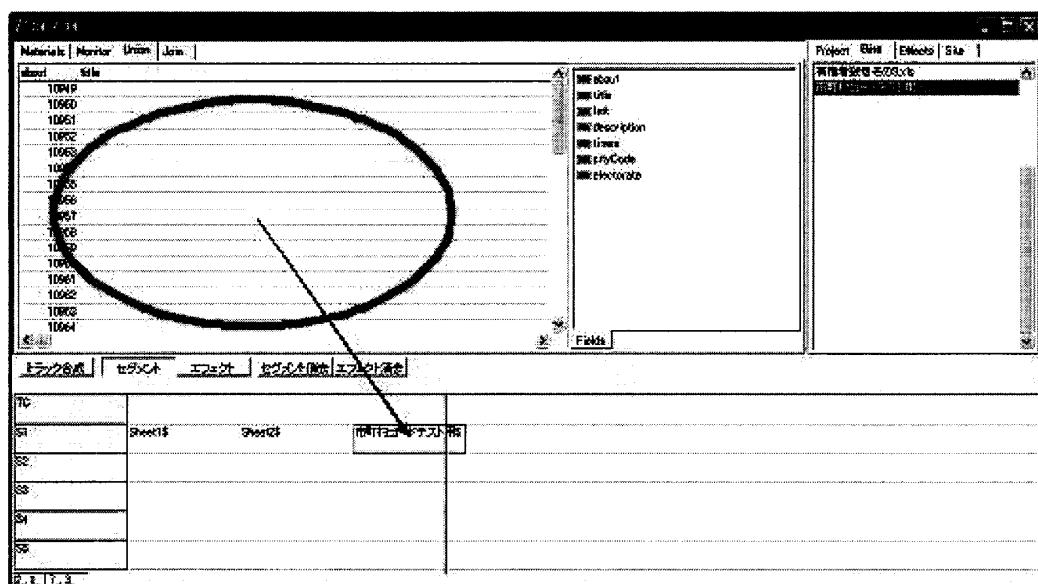


図11

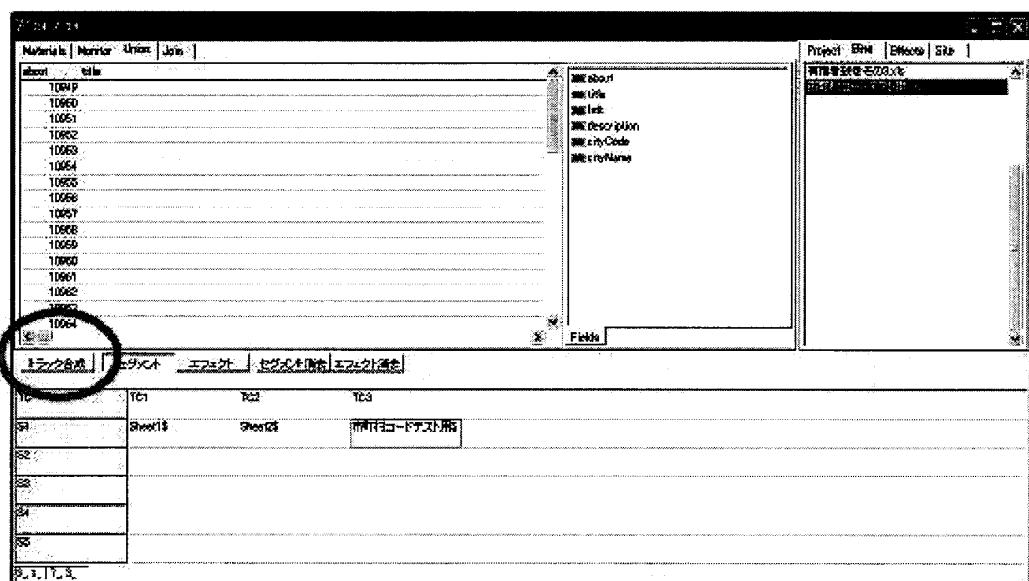


図12

12). つぎに95年参院選データを結合させるために、Joinタブを開く（図13）。すると左側上部の枠内の上下に、合成するSheetのデータの一部が表示される。ここで、中央上部に結合すべきデータ項目を入力し（この部分を選択式にすることが今後の課題の一つである）、プレビューボタンをクリックすると、左側上部の枠内の中央に、結合されたデータセットが表示される（図14）。そして中央に表示されたデータセットを右クリックし、出力を選択すると、新しいファイルが作成される。完成した出力データは図15に示される。

このように、依然として発展途上にある部分は残されているものの、操作の簡便性、視認性、ユーザーの自由度、データ入力の柔軟性をもつ、データベースシステムを構築したのである。

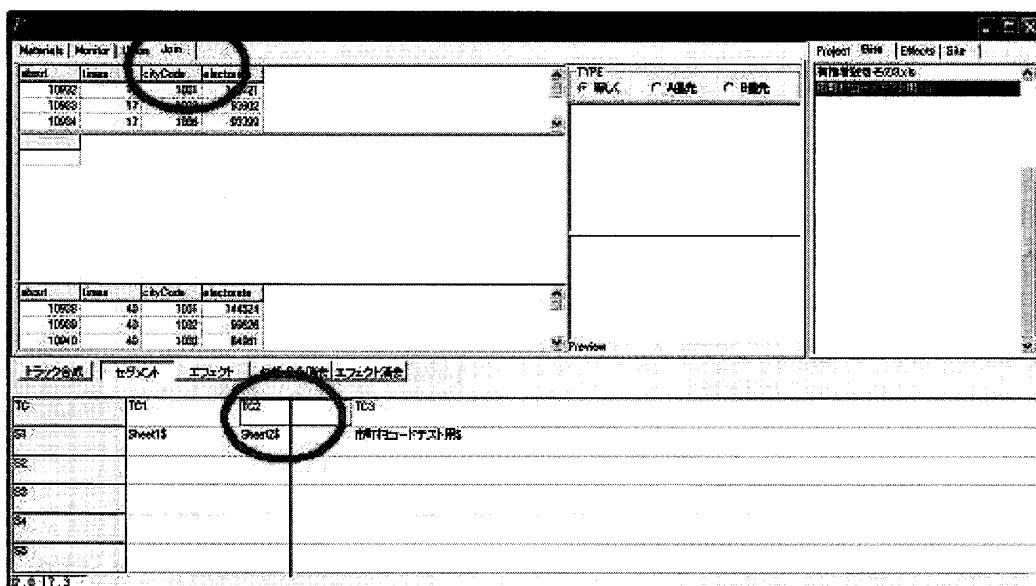


図13

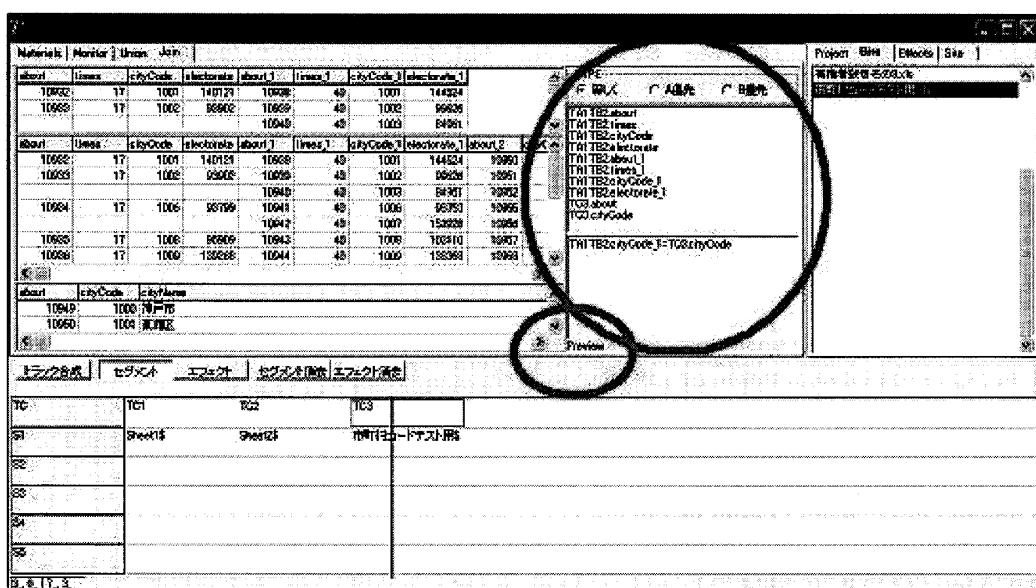


図14

about	times	cityCode	electorate	about_1	times_1	cityCode_1	electorate	about_2	cityCode_2	cityName
2 10932	17 1001	140121	10938	40 1001	144524	10950	1001	東灘区		
3 10933	17 1002	93902	10939	40 1002	99626	10951	1002	灘区		
4			10940	40 1003	84961	10952	1003	中央区		
5 10934	17 1006	93799	10941	40 1006	98753	10955	1006	兵庫区		
6			10942	40 1007	153928	10956	1007	北区		
7 10935	17 1008	95909	10943	40 1008	103110	10957	1008	長田区		
8 10936	17 1009	139268	10944	40 1009	139368	10958	1009	須磨区		
9			10945	40 1010	178594	10959	1010	垂水区		
10			10946	40 1011	129113	10960	1011	西区		
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										

図15

6 課題と展望

本論文では、日本政治に関するデータ環境の整備が不十分なことにより、研究の発展が阻害されているという問題意識から構築した、リレーションナルデータベースを紹介してきた。そこでは、映像編集ソフトから着想を得たインターフェースを採用し、操作の簡便性、視認性、ユーザーの自由度、データ入力の柔軟性をもつシステムであることを示してきた。

もちろん、本システムには、まだ改善すべき点が残されている。本文中でも述べたように、結合する項目を、入力式から選択式に修正する必要があるし、そもそもデータファイルの選択についても改良の余地は残されている。さらに、データの集計方法（合算や按分）についても、ユーザーが選択できるようなシステムにすることも必要である。また、投票率や得票率、あるいは有効候補者数など、分析において頻繁に用いられる変数については自動計算機能を備えるべきであろう。

とはいえ、こうした試みを始めたことで、日本政治データに関するリレーションナルデータベースの構築が可能であることは明らかになった。今後は、日本政治研究の発展に貢献しうるよう、更なる構築作業を進めていくことにしたい。

謝辞

本研究は、平成16、17年度関西大学特別研究・教育促進費において、研究課題「政治・行政関連データベースに関する基礎的研究」として研究費を受け、その成果を公表するものである。またシステム構築にあたっては、近藤育雄氏の多大な協力を得た。記して感謝の意を表したい。

注

- (1) 選挙に関する分析の場合、時間的コストの問題は重要である。すぐに“つぎの選挙”がやってくるからである。ところが、総務省や地方自治体からの公式データの発表までには、意外と時間がかかる。近年ではHPに結果を掲載する自治体も増えてきており、入力という行為については、非常に便利になった。ところが、HP上の“公式”データにエラーが生じているケースがある。安易に掲載データに頼ると、思わぬ落とし穴にはまる可能性がある。したがってエラーチェックには依然として慎重さが求められる。また（政治系でデータ分析を行うものにとっては“言うまでもない事”であるが）、地方自治体ごとにデータ入力のフォーマットはバラバラである。さらに言うと、同じ自治体であっても、年度によってフォーマットが異なるケースも多い。したがって、これら諸点を考慮しながら分析可能な状態へとデータの加工を進め、エラーを発見した場合には自治体に直接問い合わせるといった作業が必要になる。

忘れてはならないのが、日本では、衆院選だけでなく、3年に一度の参院選、4年に一度の知事選、都道府県議会選なども存在することである。少しでも加工の手を緩めると、すぐに別の選挙が実施されてしまうのである。このことから時間コストの問題は重要である。

- (2) これがなぜ“符合しない”のかについては、たとえば増山・山田^[5]、117ページ。
- (3) この問題は、分析に使用したデータを公開すれば、簡単に解決するように見える。しかし著作権の問題を看過してはならない。とりわけ市販されているデータを無料で公開してしまうようなことは避けねばならない。またサーバイの個票データについても、個人が特定されると困るという理由で、公開に消極的な態度を示す者が少なからず存在する。
- (4) ほとんどの論文では、データの加工過程（計算方法）が示されている。しかし、その膨大な作業を再現し、再テストされることはあるだろう。
- (5) 日本政治研究におけるデータ分析は、選挙研究を中心に発展した。その歴史的経緯は、三宅・内田^[6]に詳しい。
- (6) 水崎データや川人データは、固定長式のテキストデータであり、簡便なプログラムによって政党別や都道府県別の集計・演算も可能である。すなわち、これらのデータは“加工”的手間も軽減される優れたデータセットといえる。
- (7) 実際の分析では、得票率だけでなく、独自の変数（地盤指数）を作成した。具体的な作業としては、まず、各市町村の全投票数を、選挙区全体の投票数で除して、理論上の得票構成比を算出する。つぎに、候補者の各市町村における得票数を、選挙区全体の得票数で除して、実際の構成比を算出する。そして、実際の構成比を、理論上の構成比で除することにより算出した。
- (8) ただし中選挙区時代のファイルと並立制導入後のファイルでは、入力フォーマットが異なっている。そのため、別のプログラムを作成する必要がある。
- (9) プログラムといつても簡単なもので構わない。多少の遠回りをして、処理に不必要的時間がかかったとしても、結果として目的が果たせれば良いのである。したがって、初歩的なプログラミングの知識があれば、それで十分である。
- (10) 日経FQの場合は、このような「設定ができる」と書いた方が正確である。ユーザーは、自治体名・財政費目・年度を、自由な組み合わせで行と列に設定することができる。
- (11) 実際の問題は、同一名称の自治体だけではない。漢字の使い方ですら一様ではないからである。例えば四條畷市と表記するデータもあれば、四条畷市と表記するものもある。ケとケの違いといった

問題も生じる、場合もある。

- (12) 中選挙区制においても福島県郡山市は二つに分割され、異なる選挙区に属していた。すなわち分割市町村の問題は、小選挙区比例代表並立制特有のものではない。なおこの問題について、蒲島・甲斐^[7]は、政令指定都市にダミーの選挙区コードを付与することで対処している。しかし、他のデータとの結合を考えれば、JISコードを利用する方が後の手間が少ないようと思われる。
- (13) ただし、このような「按分」が妥当であるとは言い切れない。分析からケースを除外することが望ましいこともある。
- (14) この点は、小林^[8]でも指摘されている。小林が作成したデータベースシステムでは、この問題を克服するために独自のコードを作成し、各市区町村に付与している。

主要参考文献

- [1] 名取良太：選挙制度改革と利益誘導政治、選挙研究、17号、pp128-141（2002）。
- [2] 森裕城：得票データ分析の課題と方法—2003年総選挙の得票分析を事例として、日本選挙学会2004年度研究会報告論文（2004）。
- [3] 蒲島郁夫・綿貫讓治・三宅一郎・小林良彰・池田謙一：JES II コードブック（変動する日本人の選挙行動⑥）、p1009、木鐸社、東京（1998）。
- [4] 堀内勇作・名取良太：二大政党制を阻害する地方選挙制度、社会科学研究、第58巻第5・6号、pp21-32（2007）。
- [5] 増山幹高・山田真裕：計量政治分析入門、p175、東京大学出版会、東京（2004）。
- [6] 内田満・三宅一郎：座談会 選挙研究事始めの頃、選挙研究、15号、pp56-64（2000）。
- [7] 蒲島郁夫、甲斐康誠〔編〕：全国市町村選挙・社会経済情報：MKKデータ・コードブック、〔筑波大学多目的データ・バンク〕、（出版年不明）。
- [8] 小林良彰：選挙区分別・市区町村別・政治関連データベース、（1998）。