

Title	国立大学教授就任に係る異動・昇格に関するカテゴリカルデータ分析
Author(s)	細坪, 護挙
Citation	年次学術大会講演要旨集, 25: 193-198
Issue Date	2010-10-09
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9275
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

国立大学教授就任に係る異動・昇格に関するカテゴリカルデータ分析

○細坪 護孝（文部科学省 科学技術政策研究所）

1. はじめに—高等教育教員異動データベース(HM-DB)の構築

本研究では国立大学の教授に就任する直前直後のデータに着目して、主にその異動メカニズムなどを調べることを目的とする。2006 年からの「全国大学職員録」(廣潤社、以下「職員録」という)[1]に関する予備調査の結果[2]、職員録は政府統計と比べても高い網羅性を示し、高い質を有する研究資料であることが分かった。

紙媒体である観測時点の「全国大学職員録」を電子化し、異時点間のデータを接続した。その際の対象者は「掲載時点間において一度でも国公立大学教員になったことのある者」と定義している。

この接続の際、データクレンジング等処理・対処すべき点として、主に次の事項がある。

- ・ 氏名の漢字の表記の揺れ(ex. JIS 漢字と unicode)による接続性悪化には、辞書作成等により回避。
- ・ 大学内で同一教員が複数肩書を兼任・併任する場合には、主務と思しき職務に統一。
- ・ 同姓同名の教員を区別するためには、所属先や職位、年齢等により判別。
- ・ 結婚等により苗字を変更したと思われる教員は手作業で接続。

その他にも、時点間の接続などを実施するに当たって様々な処理を実施した[2][3][4]。

上記のデータクレンジングの大半は、接続プログラムにより科学的・自動的に実施した。こうして、高等教育教員異動データベース(Higher-education teachers' Mobility Data Base: 略称 HM-DB)を構築した。

本研究で使用した HM-DB では、職員録の観測時点として、国公立大学 8 時点(1988, 1991, 1994, 1997, 2000, 2003, 2005, 2006 年)・私立大学 7 時点(1988, 1991, 1994, 1997, 2000, 2003, 2006 年)に加えて、2008 年の「文部科学省・国立大学法人等職員録」(財団法人文教協会)における国立大学の情報を使用した[5]。

2. 職員録と政府統計における大学教員の比較

2.1 職位別・観測時点別比較

HM-DB の教員属性別・観測時点別の掲載情報の偏りを確認する必要がある。そのために、信頼性の高いデータとして、文部科学省が実施する基幹統計である学校基本調査の調査結果を活用して、職位別・観測時点別に比較した。ここで、学校基本調査を信頼性の高いデータとし、職員録による HM-DB の計数値を観測値と考え、各観測時点の計数値に対して自由度 3(2008 年のみ 4)のカイ二乗検定を使用できる。この検定結果から、全ての観測時点において、学校基本調査と HM-DB の構成には差があることが分かった(0.1%有意)。つまり、HM-DB には全ての観測時点で助手や講師の人数が少ない。この背景には、もともと

職員録は専門家を検索するための冊子として出版・販売されてきた経緯があると考えられる。

次に、それぞれの職位について観測時点別の変動を上記と同様の方法で調べた。検定の結果、教授($P=0.303$)、助教授・准教授($P=0.424$)、及び講師($P=0.621$)の3職位では観測時点によって掲載率は変わらない。一方、助手・助教は異なり、観測時点によるばらつきが大きい(0.1%有意)。したがって、若手研究者が多いと思われる助手・助教に関するデータの取り扱いには慎重を期する必要がある。

2.2 年齢階層別・観測時点別比較

次に、年齢階層と観測時点の関係を調べる。国立大学教員の年齢階層の構成については、文部科学省の基幹統計である学校教員統計調査の内外挿した推計値を信頼性の高いデータとする。このデータと、職員録の HM-DB による年齢階層別・観測時点別の国立大学教員数を比較する。カイ二乗検定を実施すると、各観測時点における両者の年齢階層の構成は異なることが分かる(0.1%有意)。一方、各年齢階層における観測時点の効果を調べたところ、この場合も両者の構成は異なることが分かる(0.1%有意)。

3. 国立大学の4群分類

3.1 コンセプト

本研究では、86 の国立大学(2009 年 11 月時点)を当所における論文分析の結果[6]を活用して、4つの群、即ち、第一群(G1)、第二群(G2)、第三群(G3)、第四群(G4)に分けた(附録参照)。これは大学のアウトプットを重視した分類方法である。

3.2 群別教授総数の推移・比較

次の課題はこの4群の教授数が HM-DB において適切に把握されているかどうかを調べることである。そこで、2006 年 10 月から当所が実施している「大学等における科学技術・学術活動実態調査報告」(以下「大学把握調査」という)[7][8]の掲載情報—具体的には公表されている 2005 年と 2006 年の教授数—を信頼性の高いデータとして活用する。

G1・G2 教授数に関して、HM-DB と大学把握調査間で同時期における構成割合のカイ二乗検定を行うと、それぞれの観測時点の HM-DB の G1・G2 教授数は大学把握調査を適切に反映することが分かる($G1:P_{2005}=0.797$, $P_{2006}=0.625$, $G2:P_{2005}=0.208$, $P_{2006}=0.444$)。

一方、同様に G3 に対してカイ二乗検定を行うと、 $P_{2005}=0.000$, $P_{2006}=0.001$ となり、両者に乖離が見られる。これは、同群の新潟大学と熊本大学の教員データが職員録に掲載されていない観測時点があることが原因である。企業が実施する職員録への協力は任意であるため、両大学の責ではない。両大学を除いた場合、カイ二乗検定統計量は大幅に減少し、 $P_{2005}=0.999$, $P_{2006}=1.000$ となる。

最後に、G4 に対してカイ二乗検定を行うと、 $P_{2005}=0.940$, $P_{2006}=0.997$ となり、それぞれの観測時点の HM-DB の G4

教授数は大学把握調査を適切に反映することが分かる。しかし、G4の構成大学を見ると、HM-DBにおける大分大学、宮崎大学の教授数の把握は十分ではない。G3とG4の検定結果の違いは構成大学数、即ち自由度に起因する。

全観測時点を通じた各群内の教授総数の傾向は増加傾向にある。その単回帰の傾きの有意性検定では、全ての群で0.1%有意である。

4. 3 因子クロス表分析

本研究が目的とする国立大学4群への教授就任状況を調べるためには、以下の事項に関する観測値を把握し、統合的に検討する必要がある。即ち：①教授就任先(4群の国立大学)、②教授就任元機関(4群の国立大学)、③出身機関(4群の国立大学)、④観測時点(9時点による8区間)、⑤教授就任時の昇格の有無、の5因子を使用する。

他にも、教授就任者の年齢・世代、専門分野、異動距離、大学の所在地等様々な因子が考えられるが、年齢や世代については2. 2で現時点の分析には問題があることが判明している。他の因子に関しては、無暗な因子数増加による説明力過剰、説明因子の優先順位の問題から、今後の課題とする。

本章では、冒頭に設定した5因子のうち、いくつかの変量に対して時系列のクロス表を構成し、分析する。直観的には就任先機関に対してまず影響を及ぼしそうなのは就任元機関、又は出身機関と考えられる。その観点から、以下の2つのケースを考える。

- ②就任元機関×①就任先機関×④観測時点
- ③出身機関×①就任先機関×④観測時点

4. 1 「②就任元機関×①就任先機関×④観測時点」のクロス表

まず、上記1)のクロス表を表1に示す。

1988-1991(3年間平均値)						1991-1994(3年間平均値)					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	175	22	16	19	231	G1	203	16	15	23	257
G2	16	255	13	17	301	G2	271	9	27	328	
G3	5	9	207	8	230	G3	4	13	204	13	234
G4	9	14	8	318	349	G4	8	15	10	317	350
総計	205	300	244	362	1,111	総計	236	316	238	380	1,169

1994-1997(3年間平均値)						1997-2000(3年間平均値)					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	211	18	20	26	275	G1	180	23	12	24	238
G2	20	277	15	17	329	G2	18	249	10	16	293
G3	5	16	219	14	253	G3	9	14	187	12	222
G4	15	17	8	322	362	G4	11	16	12	297	336
総計	251	327	263	378	1,220	総計	218	301	221	350	1,090

2000-2003(3年間平均値)						2003-2005(2年間平均値)					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	173	18	14	13	219	G1	179	20	9	16	224
G2	20	229	12	14	275	G2	22	248	10	11	290
G3	10	12	172	12	207	G3	6	11	180	10	206
G4	11	18	9	287	325	G4	11	19	10	291	331
総計	214	278	207	327	1,026	総計	218	297	208	327	1,049

2005-2006(1年間観測値)						2006-2008(2年間平均値)					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	177	18	3	16	214	G1	210	13	13	13	249
G2	20	224	15	11	270	G2	22	215	14	16	266
G3	8	7	142	9	166	G3	6	6	251	9	271
G4	2	22	6	293	323	G4	6	15	15	336	372
総計	207	271	166	329	973	総計	243	248	293	374	1,157

(注) HM-DB から作成

表1 ②就任元機関×①就任先機関×④観測時点のクロス表

表1の各観測時点について、カイ二乗統計量を使用した独立性検定を実施したところ、全ての観測時点で就任先と就任元との独立性の帰無仮説は棄却される(0.1%有意)。表1の観測値から、対角成分、特に自校昇格者数が非常に大きいと考えられる。このことは調整残差分析からも

確認される。そのため、表1から自校昇格者を除いたクロス表を分析した。

この各観測時点について、カイ二乗統計量を使用した独立性検定を実施したところ、2005-2006年の観測時点以外では就任先と就任元との独立性の帰無仮説は棄却されない($P_{1988-1991}=0.719$, $P_{1991-1994}=0.223$, $P_{1994-1997}=0.246$, $P_{1997-2000}=0.922$, $P_{2000-2003}=0.746$, $P_{2003-2005}=0.650$, $P_{2005-2006}=0.006$, $P_{2006-2008}=0.305$)。即ち、自校昇格者を除けば、ほとんどの観測時点において異動元と異動先の国立大学群間に特別な関係は存在しない。

それでは、どの群間の異動が特徴的なのか。調整残差分析[9], [10], [11]を行うと、全ての観測時点においてG2からG1への異動教授が他の群間の異動よりも多い。逆方向の場合はそうではない。あとは、2005年以降、G4からG1への異動教授は他の群間よりも少ない。G2からG4も強い傾向ではないが、減少している。これまでの分析結果とこの調整残差分析の結果から、G2からG1教授就任するケースを除き、他の群間の異動教授は少なく、「異動の群内ブロック化」の傾向が強くなっている。

その証拠の一つとして、自校昇格者を除く表1の各観測時点の対角成分の和を総度数で割ったものの推移(表2)は、明確な増加傾向を示す。図表の右端のP値とは、この数値を単回帰した際の傾きの有意性のP値である。

	1988-1991	1991-1994	1994-1997	1997-2000	2000-2003	2003-2005	2005-2006	2006-2008	参考 P 値
対角成分の和/総度数	0.220	0.220	0.220	0.244	0.248	0.242	0.271	0.285	0.0027

表2 自校昇格者を除いた②就任元機関×①就任先機関×④観測時点のクロス表の対角成分の和を総度数で割ったものの推移

4. 2 「③出身機関×①就任先機関×④観測時点」のクロス表

次に、冒頭の2)に示した③出身機関×①就任先機関×④観測時点のクロス表(表3)を分析する。

1988-1991						1991-1994					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	189	116	96	144	545	G1	216	112	100	155	583
G2	14	172	87	128	401	G2	17	189	78	140	423
G3	1	5	49	12	67	G3	2	6	49	13	70
G4	1	6	12	78	97	G4	1	9	11	72	94
総計	205	300	244	362	1,111	総計	236	316	238	380	1,169

1994-1997						1997-2000					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	232	123	107	147	610	G1	192	108	81	136	517
G2	17	190	88	135	430	G2	23	178	71	123	395
G3	1	6	53	16	76	G3	2	7	54	15	78
G4	1	9	14	80	104	G4	1	8	16	75	100
総計	251	327	263	378	1,220	総計	218	301	221	350	1,090

2000-2003						2003-2005					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	195	102	74	117	488	G1	196	107	76	124	502
G2	14	158	75	113	360	G2	15	171	70	116	372
G3	2	8	44	14	68	G3	2	5	48	12	66
G4	3	9	15	83	110	G4	5	14	16	76	110
総計	214	278	207	327	1,026	総計	218	297	208	327	1,049

2005-2006						2006-2008					
G1	G2	G3	G4	総計		G1	G2	G3	G4	総計	
G1	171	97	53	129	450	G1	209	72	100	113	493
G2	29	155	68	109	361	G2	24	159	108	132	423
G3	5	10	31	13	59	G3	5	7	67	21	100
G4	2	9	14	78	103	G4	5	11	18	108	142
総計	207	271	166	329	973	総計	243	248	293	374	1,157

(注) HM-DB から作成

表3 ③出身機関×①就任先機関×④観測時点のクロス表

これに対して、カイ二乗統計量を用いた就任先と出身との独立性に関する検定を実施すると、全ての観測時点において独立性の帰無仮説は棄却される(0.1%有意)。それは表3が、対角成分の右上側の度数の大きい、偏りの大きなクロス

表となっていることから明らかである。これは、出身者の多くが G1 又は G2 に集中していることを意味する。

就任元の分析の場合と同様に、出身元に関しても対角成分から自校昇格者を除いたクロス表を分析した。これに対して、カイ二乗統計量を用いた就任先と出身群との独立性に関する検定を実施すると、全ての観測時点において独立性の帰無仮説は棄却される(1%有意)。つまり、自校昇格者を除いても、全ての時点で出身と就任先は独立ではない。

4.3 クロス表における G3・G4 カテゴリー

前述の分析の元となるクロス表(「②就任元機関×①就任先機関×④観測時点」の表 1、並びに「③出身機関×①就任先機関×④観測時点」の表 3)から、G3 と G4 が類似していることに気付く。もし、この2群が統計学的に同じ挙動を示していれば、2つの群に分ける必要はなくなる。

ある2つの群と、他群(G1・G2)との関係が同じであれば、その群は同一群と見なす。そう考えれば、表 1 及び表 3 において、各観測時点における、右上の2×2セル、左下の2×2セルが独立であれば、セルを統合できる。

こうして、まずは表 1 の「②就任元機関×①就任先機関×④観測時点」に対してフィッシャーの正確確率検定を行ったところ、2005-2006 年の観測時点において右上・左下ともに独立性の帰無仮説が棄却される(独立ではない)。一方、表 3 の「③出身機関×①就任先機関×④観測時点」に対してフィッシャーの正確確率検定を行ったところ、全ての観測時点における、右上2×2セル及び左下2×2セルの独立性の帰無仮説が棄却されない(独立である)。

以上の結果から、表 3 の「③出身機関×①就任先機関×④観測時点」では、G3 と G4 を統合することができると考えられる。

そこで表 3 の G3 と G4 を統合し、NG3 と呼称して分析する。まず、統合したクロス表を示すと表 4 となる。

1988-1991					1991-1994				
	G1	G2	NG3	総計		G1	G2	NG3	総計
G1	189	116	240	545	G1	216	112	255	583
G2	14	172	215	401	G2	17	189	218	423
NG3	2	11	151	164	G3	3	15	146	164
総計	205	300	606	1,111	総計	236	316	618	1,169

1994-1997					1997-2000				
	G1	G2	NG3	総計		G1	G2	NG3	総計
G1	232	123	255	610	G1	192	108	217	517
G2	17	190	223	430	G2	23	178	194	395
NG3	2	15	163	180	G3	3	15	160	178
総計	251	327	641	1,220	総計	218	301	571	1,090

2000-2003					2003-2005				
	G1	G2	NG3	総計		G1	G2	NG3	総計
G1	195	102	191	488	G1	196	107	200	502
G2	14	158	188	360	G2	15	171	186	372
NG3	5	17	156	178	G3	7	19	151	176
総計	214	278	534	1,026	総計	218	297	535	1,049

2005-2006					2006-2008				
	G1	G2	NG3	総計		G1	G2	NG3	総計
G1	171	97	182	450	G1	209	72	212	493
G2	29	155	177	361	G2	24	159	240	423
NG3	7	19	136	162	G3	10	18	214	242
総計	207	271	495	973	総計	243	248	666	1,157

(注) 表 3 から作成

表 4 第三群と第四群を統合した「③出身機関×①就任先機関×④観測時点」のクロス表

表 4 は、表 3 を縮約したものである。ここでは別のアプローチを試みる。「出身機関の群から、より上位群の大学教授に就任する割合は、どの出身機関の群でも同じである」を帰無仮説とし仮説検定を考える。

上記の帰無仮説を調べるためには、表 4 の網掛部分に示された部分の行和に対する割合が各観測時点において等

しいとした場合を理論値として、観測値との差に対してカイ二乗検定すればよい。

G2 出身で G1 への教授就任者数を n_{21}^i 、NG3 出身で G1 への教授就任者数を n_{31}^i 、NG3 出身で G2 への教授就任者数を n_{32}^i と表すと、上記のカイ二乗検定統計量は、次式になる。

$$X^2 = \frac{\left\{ \frac{n_{\cdot 2}^i}{3} \times \left(\frac{n_{21}^i}{n_{\cdot 2}^i} + \frac{n_{31}^i}{n_{\cdot 3}^i} + \frac{n_{32}^i}{n_{\cdot 3}^i} \right) - n_{21}^i \right\}^2}{\frac{n_{\cdot 2}^i}{3} \times \left(\frac{n_{21}^i}{n_{\cdot 2}^i} + \frac{n_{31}^i}{n_{\cdot 3}^i} + \frac{n_{32}^i}{n_{\cdot 3}^i} \right)} + \frac{\left\{ \frac{n_{\cdot 3}^i}{3} \times \left(\frac{n_{21}^i}{n_{\cdot 2}^i} + \frac{n_{31}^i}{n_{\cdot 3}^i} + \frac{n_{32}^i}{n_{\cdot 3}^i} \right) - n_{31}^i \right\}^2}{\frac{n_{\cdot 3}^i}{3} \times \left(\frac{n_{21}^i}{n_{\cdot 2}^i} + \frac{n_{31}^i}{n_{\cdot 3}^i} + \frac{n_{32}^i}{n_{\cdot 3}^i} \right)} + \frac{\left\{ \frac{n_{\cdot 3}^i}{3} \times \left(\frac{n_{21}^i}{n_{\cdot 2}^i} + \frac{n_{31}^i}{n_{\cdot 3}^i} + \frac{n_{32}^i}{n_{\cdot 3}^i} \right) - n_{32}^i \right\}^2}{\frac{n_{\cdot 3}^i}{3} \times \left(\frac{n_{21}^i}{n_{\cdot 2}^i} + \frac{n_{31}^i}{n_{\cdot 3}^i} + \frac{n_{32}^i}{n_{\cdot 3}^i} \right)}$$

全ての観測時点について調べたところ、就任先と出身との独立性の帰無仮説は棄却されない。即ち、それぞれの観測時点において、G2 出身者が G1 の教授に就任する割合、NG3 出身者が G1 の教授に就任する割合、同じく NG3 出身者が G2 の教授に就任する割合は等しい。加えて、フィッシャーの正確確率検定を行ったところ、1997-2000 年、2006-2008 年の2つの観測時点のみ、5%有意水準で就任先と出身群との独立性を確認できた。

なお、観測時点に応じてこの割合は変化する。この出身機関から上位の就任先機関への教授就任割合は年々増加している(参考 P 値=0.02)。

逆に、就任先機関別の構成割合から、「ある出身機関から上位の就任先機関の教授に就任者の合計の割合は、どの就任先機関の群でも同じである」を帰無仮説とする仮説検定を考える。この帰無仮説を調べるためには、表 4 の網掛部分に示された割合の列和が各観測時点において等しいとした場合を理論値として、観測値との差に対してカイ二乗検定すればよい。

G2 出身で G1 への教授就任者数を n_{21}^i 、NG3 出身で G1 への教授就任者数を n_{31}^i 、NG3 出身で G2 への教授就任者数を n_{32}^i と表すと、上記のカイ二乗検定統計量は、次式となる。

$$X^2 = \frac{\left\{ \frac{n_{\cdot 1}^i}{2} \times \left(\frac{n_{21}^i}{n_{\cdot 1}^i} + \frac{n_{31}^i}{n_{\cdot 2}^i} + \frac{n_{32}^i}{n_{\cdot 2}^i} \right) - (n_{21}^i + n_{31}^i) \right\}^2}{\frac{n_{\cdot 1}^i}{2} \times \left(\frac{n_{21}^i}{n_{\cdot 1}^i} + \frac{n_{31}^i}{n_{\cdot 2}^i} + \frac{n_{32}^i}{n_{\cdot 2}^i} \right)} + \frac{\left\{ \frac{n_{\cdot 2}^i}{2} \times \left(\frac{n_{21}^i}{n_{\cdot 1}^i} + \frac{n_{31}^i}{n_{\cdot 2}^i} + \frac{n_{32}^i}{n_{\cdot 2}^i} \right) - n_{32}^i \right\}^2}{\frac{n_{\cdot 2}^i}{2} \times \left(\frac{n_{21}^i}{n_{\cdot 1}^i} + \frac{n_{31}^i}{n_{\cdot 2}^i} + \frac{n_{32}^i}{n_{\cdot 2}^i} \right)}$$

全ての観測時点を上式で調べたところ、2005-2006 年を除き、就任先群と出身群との独立性の帰無仮説は棄却されない。この結果はフィッシャーの正確確率検定でも同じだった。即ち、2005-2006 年を除いた観測時点において、G1 大学教授就任における G2 及び NG3 出身者の割合は、G2 大学教授就任における NG3 出身者の割合と等しい。

また、この場合も観測時点により割合が変化する。この就任先機関に対する下位の群の出身機関からの教授就任割合は年々増加している(参考 P=0.01)。

加えて、就任先機関別の構成割合から、「ある出身機関群から直近の上位群の就任先機関教授に就任する者の人数の割合は、どの就任先機関の群でも同じである」を帰無仮説とする仮説検定を考える。この帰無仮説を調べるためには、表 4 の斜線部分に示された部分の列和に対する割合が各観測時点で等しいとして理論値を計算し、観測値との差に対してカイ二乗検定すればよい。

G2 出身で G1 への教授就任者数を n'_{21} 、新 G3 出身で G2 への教授就任者数を n'_{32} と表すと、上記のカイ二乗検定統計量は次式となる。

$$X^2 = \frac{\left\{ \frac{n'_1}{2} \times \left(\frac{n'_{21}}{n'_1} + \frac{n'_{32}}{n'_2} \right) - n'_{21} \right\}^2}{\frac{n'_1}{2} \times \left(\frac{n'_{21}}{n'_1} + \frac{n'_{32}}{n'_2} \right)} + \frac{\left\{ \frac{n'_2}{2} \times \left(\frac{n'_{21}}{n'_1} + \frac{n'_{32}}{n'_2} \right) - n'_{32} \right\}^2}{\frac{n'_2}{2} \times \left(\frac{n'_{21}}{n'_1} + \frac{n'_{32}}{n'_2} \right)}$$

上式で全ての観測時点を調べたところ、全ての観測時点において就任先群と出身群との独立性の帰無仮説は棄却されない。この結果はフィッシャーの正確確率検定でも同じだった。即ち、全ての観測時点において、G1 教授就任における G2 出身者の割合は、G2 教授就任における NG3 出身者の割合と等しい。

また、この場合でも観測時点により割合は変化する。この就任先機関に対する直近下位の出身機関群からの教授就任割合は年々増加している(参考 P=0.03)。

5. 5 因子対数線形モデル

4. の結論から、就任元における自校昇格者、出身機関における自校出身者の効果は大きく、別カテゴリーに分類することが適切であることが判明した。そこで、自校昇格者・自校出身者を加えた表 5(就任先と就任元)及び表 6(就任先と出身)に対して独立性の検定を実施すると、就任元機関に自校のカテゴリーを追加した表 5 では、2005-2006 年の 1 点を除く全ての観測時点で就任先と就任元との独立性の帰無仮説が棄却されない(独立である。 $P_{1988-1991}=0.840$, $P_{1991-1994}=0.283$, $P_{1994-1997}=0.378$, $P_{1997-2000}=0.921$, $P_{2000-2003}=0.463$, $P_{2003-2005}=0.209$, $P_{2005-2006}=0.001$, $P_{2006-2008}=0.152$)。一方、出身機関に自校出身のカテゴリーを追加した表 6 では全ての観測時点で就任先と出身との独立性の帰無仮説が棄却される(独立ではない。全ての観測時点で $P=0.000$)。

これは、同じ「自校」といっても意味が異なるからであり、内部昇格者が多いこと、G3 や G4 の教授就任者の多くが G1 又は G2 の出身であることなど、就任元と出身元との構造の大きな違いによる。

1988-1991						1991-1994					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校	166	242	202	301	911	194	256	199	297	947	
G1	9	22	16	19	65	9	16	15	23	63	
G2	16	13	13	17	59	21	15	9	27	72	
G3	5	9	6	8	28	4	13	5	13	35	
G4	9	14	8	17	48	8	15	10	20	53	
総計	205	300	244	362	1,111	236	316	238	380	1,169	

1994-1997						1997-2000					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校	195	260	212	306	974	167	232	179	278	856	
G1	15	18	20	26	80	13	23	12	24	72	
G2	20	16	15	17	69	18	16	10	16	61	
G3	5	16	7	14	41	9	14	8	12	43	
G4	15	17	8	15	56	11	16	12	20	58	
総計	251	327	263	378	1,220	218	301	221	350	1,090	

2000-2003						2003-2005					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校	163	209	164	270	805	168	230	173	278	849	
G1	10	18	14	13	56	11	20	9	16	56	
G2	20	20	12	14	67	22	18	10	11	60	
G3	10	12	8	12	43	6	11	7	10	33	
G4	11	18	9	17	55	11	19	10	13	53	
総計	214	278	207	327	1,026	218	297	208	327	1,049	

2005-2006						2006-2008					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校	161	206	137	281	785	191	201	241	319	952	
G1	16	18	3	16	53	19	13	13	13	58	
G2	20	18	15	11	64	22	14	14	16	65	
G3	8	7	5	9	29	6	6	10	9	30	
G4	2	22	6	12	42	6	15	15	17	53	
総計	207	271	166	329	973	243	248	293	374	1,157	

(注) HM-DB から作成。自校昇格者を別のカテゴリーに設置

表 5 ②就任元機関×①就任先機関×④観測時点のクロス表

1988-1991						1991-1994					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校出身	164	137	45	50	396	175	148	43	43	409	
G1	25	116	96	144	381	41	112	100	155	408	
G2	14	35	87	128	264	17	40	78	140	275	
G3	1	5	4	12	22	2	6	7	13	27	
G4	1	6	12	28	47	1	9	11	30	51	
総計	205	300	244	362	1,111	236	316	238	380	1,169	

1994-1997						1997-2000					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校出身	184	155	45	45	429	160	134	46	45	384	
G1	49	123	107	147	426	32	108	81	136	357	
G2	17	36	88	135	276	23	44	71	123	261	
G3	1	6	8	16	30	2	7	8	15	32	
G4	1	9	14	35	59	1	8	16	31	56	
総計	251	327	263	378	1,220	218	301	221	350	1,090	

2000-2003						2003-2005					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校出身	154	113	36	49	352	152	128	41	44	364	
G1	41	102	74	117	333	44	107	76	124	350	
G2	14	45	75	113	247	15	44	70	116	244	
G3	2	8	8	14	32	2	5	7	12	25	
G4	3	9	15	35	62	5	14	16	32	66	
総計	214	278	207	327	1,026	218	297	208	327	1,049	

2005-2006						2006-2008					
	G1	G2	G3	G4	総計	G1	G2	G3	G4	総計	
自校出身	134	101	28	44	307	163	122	60	72	416	
G1	37	97	53	129	316	47	72	100	113	330	
G2	29	54	68	109	260	24	37	108	132	301	
G3	5	10	3	13	31	5	7	8	21	41	
G4	2	9	14	34	59	5	11	18	36	70	
総計	207	271	166	329	973	243	248	293	374	1,157	

(注) HM-DB から作成。自校出身者を別のカテゴリーに設置

表 6 ③出身機関×①就任先機関×④観測時点のクロス表

これから、これらのデータを一括して、対数線形モデルで扱う。以下の対数線形モデルの計算では、冗長性解決の手段としては、Birch の制約式 $\sum_i \lambda_{行i} = 0$ 等を使用する。これは、オッズ比の比較分析を行うためである。

まず、モデルの計算に使用する因子とそのカテゴリーを整理する。4. 冒頭に示した因子全てを使用する。4. における分析と異なるのは②のカテゴリーに対する公立大学、私立大学の追加と、因子「昇格の有無」の追加である。

本研究で取り扱うモデルでは因子数は5と比較的小さく設計し、飽和モデルから独立モデルまで階層構造を形成する約3,000式を網羅的に計算して個々のAICを比較し、最適モデルを導出した。同時に、最適モデルに次いでAICが小さく、

かつ最適モデルと階層関係を有するとともに、最適モデルよりも簡潔なモデル(この場合、敢えて複雑なモデルを採用する理由はない)に対して ΔG^2 (対数尤度比の差)及び自由度の差を計算し、カイ二乗検定によりモデル間の差の有意性を調べた。検定の結果、最適モデルと次点解間には有意な差があり、最適モデルの採択が最も合理的であると判断された。

ここで結果的に得られた最適モデルでは最高次が2次交互作用項であり、オッズ比による解釈が可能である。以後、この最適モデルを対象に検討する。4. 冒頭の標記を使用して、この最適モデルを示すと次式となる。

$$\ln(\mu_{ijklm}) = \lambda_{\phi} + \lambda_{(1)(i)} + \lambda_{(2)(j)} + \lambda_{(3)(k)} + \lambda_{(4)(l)} + \lambda_{(5)(m)} + \lambda_{(1)(2)(i,j)} + \lambda_{(1)(3)(i,k)} + \lambda_{(1)(5)(i,m)} + \lambda_{(2)(3)(j,k)} + \lambda_{(2)(5)(j,m)} + \lambda_{(3)(5)(j,m)} \dots \dots (1)$$

(1)式で非常に興味深いのは、④の観測時点だけが他のどの項とも交絡しないことである。つまり、観測時点は他の項の影響を受けずに変動する。逆に観測時点以外の項は他の全ての項との間に2次交互作用項を持つ。つまり、因子①、②、③及び⑤は互いに密接な関係にある。

この最適モデルによる元のデータの再現性を調べる。最適モデルでも、再現性が悪ければ使い物にならない。この再現性を調べるための指標として、モデルの当てはめ値と観測値との近さを要約する以下の非類似性指標 (dissimilarity index) Dを用いる[12]。

$$D = \sum_{ijklm} \left| n_{ijklm} - \hat{\mu}_{ijklm} \right| / (2 \sum_{ijklm} n_{ijklm}) \dots \dots (2)$$

この指標は0~1の値をとり、値が小さいほど当てはまりが良い。(1)式の最適モデルでは、(2)式はD=0.087となり、最適モデルは観測値の91%以上を説明していると考えられ、社会科学のデータとしては妥当と評価できる。

この場合、対数線形モデルには構造的ゼロを包含している。具体的には、「②就任元機関」が自校、「⑤昇格の有無」が昇格なし、となっているセルは教授に就任しておらず、構造的にゼロとなる。対数線形モデルでは構造的ゼロのセルは対象外であるため、モデルの推定における理論上の困難は生じない[13]。

対数線形モデルのパラメータの解釈の際には、そのモデルの最高次の項に注目するものであり、高次の交互作用項を有する因子の主効果のみによる解釈は意味をなさない[12]。(1)式において、主効果として意味があるのは観測時点の効果のみである。

次に、最適モデルにおける2次交互作用項に関するオッズ比を計算する。

出身機関と就任元機関との間のオッズ比を調べた結果、出身機関と就任元機関の群が同一である場合のオッズ比が高く、その傾向は、G1・G3、G3・G4、G2・G3、G1・G4間の順に顕著である。

出身機関及び就任先機関間のオッズ比を調べると、オッズ比の高いケースの出身機関には必ず自校出身が、就任先機関にはG1が含まれている。このことは、自校出身者がG1に就任する効果が非常に大きいことを示す。

就任元機関及び就任先機関に関するオッズ比には大きな差はない。

就任元機関及び昇格の有無に関するオッズ比を調べる

と、就任元機関が G1、G2、G3、公立大学、G4、私立大学の順に昇格を伴う割合が高い。

就任先機関及び昇格の有無に関するオッズ比を調べると、就任先機関が G3、G4、G2、G1 の順に昇格を伴う割合が高いことが分かる。

最後に、出身機関及び昇格の有無に関するオッズ比を調べると、出身機関が G3、G4、G2、自校出身、G1 の順に昇格を伴う割合が高いことが分かる。

以上の分析結果を簡単に理解するため、各群の教授就任に関する典型的なキャリアパスのイメージを図1に示す。就任先の各群に対する就任元機関の上位3群を示し、それとともに、その群に対応する出身機関の上位1群を表したものである。

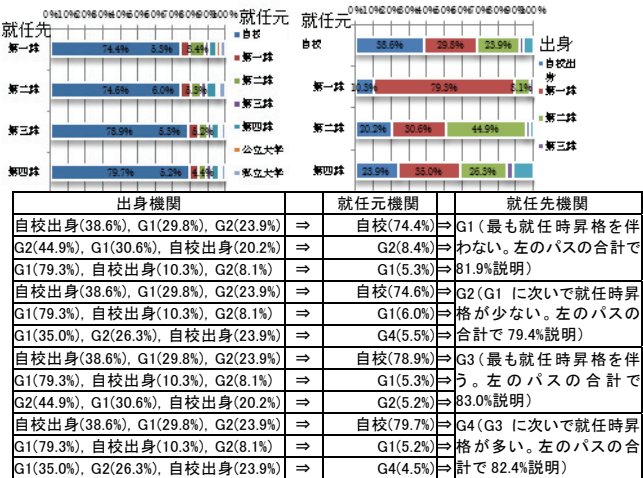


図1 対数線形モデルの当てはめ値による各群の教授就任に対する典型的なキャリアパスのイメージ

対数線形モデル分析の最後に残差分析を行う。前述の(2)式の最適モデルで説明されていない残りの1割足らずの情報調べる。5 因子の観測値と対数線形モデルによる最適モデルの当てはめ値との標準残差を求めた。最適モデルによる当てはめ値と観測値との外れ値(有意水準両側 5%に相当する絶対値 1.96 より大きいセル)の分布の法則性を調べるため、外れ値の分布を調べた(表7)。

就任先機関	G1	G2	G3	G4
外れ値数	23	15	18	8

就任元機関	自校	G1	G2	G3	G4	公立大学	私立大学
外れ値数	8	13	7	9	8	9	10

出身機関	自校出身	G1	G2	G3	G4
外れ値数	14	7	11	17	15

昇格の有無	昇格あり	昇格なし
外れ値数	29	35

観測時点	1988-1991	1991-1994	1994-1997	1997-2000	2000-2003	2003-2005	2005-2006	2006-2008
外れ値数	0	4	4	7	10	9	15	15

表7 最適モデルによる当てはめ値と観測値との外れ値(有意水準両側 5%に相当する絶対値 1.96 より大きいセル)の分布

表7に対して、就任先機関、就任元機関、出身機関及び昇格の有無と外れ値の分布に対して、外れ値数とその平均値に対してカイ二乗統計量による独立性検定を行った結果、帰無仮説は棄却されず、特段の傾向は見付からなかった。

唯一、観測時点と外れ値数との関係では独立性の帰無仮説は棄却され、観測時点が最近になればなるほど外れ値数が増えている。この傾向に関しては、誤差の正規性や均一分散が保障されない条件下での正規単回帰モデルを計算した結果でも有意な傾向を示す($P=0.00$)。

以上の結果から、元の観測値に対して、対数線形モデルの最適モデルで再現されない約1割のばらつきは、最近の観測時点に多く見られる。この問題を緩和するために、対数線形モデルの当てはめのための観測値をある観測時点を境に2つに分割することが考えられる。そこで、データを分割して対数線形モデルで分析したところ、観測時点の前半部分(-2000年若しくは-2005年の2パターン)、つまり過去の部分では(1)式と同じモデルが採択される。一方、後半の、つまり最近の部分(2000年-若しくは2005年-)ではそれぞれ(1)式に①就任先機関と④観測時点の2次交互作用項が追加されるのみであり、4.の現象はこのモデルからでは説明できない。更に、観測時点分割しない(1)式の最適モデルのAICは-3,173である一方、2000年で分割したパターンの2つの最適モデルのAICの和は-3,147、同じく2005年で分割したパターンのAICの和は-3,160となり、この3つのケースのうち観測時点分割しない場合のAICが最も小さい。以上の結果から、観測時点分割する必要性はないと考えられる。

また、観測時点に関して、近年ほど観測時点の間隔が狭まっているため、モデルからの外れ値が出やすいとも考えられる。

次に、5因子分析でも4.で観測された観測時点に依存する現象が観測されるのかどうかを調べるため、観測値と対数線形モデルの最適モデルによる当てはめ値との残差を調べる。まず、②就任元機関×①就任先機関×④観測時点の自校昇格を除く対角成分の観測時点による変化を調べたところ、残差から異動のブロック化現象が確認された(参考 $P=0.01$)。この結果から、異動のブロック化現象は対数線形モデルで説明されない約9%のばらつきに含まれていると考えられる。また、③出身機関×①就任先機関×④観測時点の就任先群の直下の群からの出身者数の変化も調べたところ、残差から就任者に対する直下の群からの出身者数の増加傾向が確認された(参考 $P_{G2 \rightarrow G1}=0.07$ 、参考 $P_{NG3 \rightarrow G2}=0.01$)。この結果から、就任者に対する直下の群からの出身者数の増加傾向も対数線形モデルで説明されない約9%のばらつきに含まれていると考えられる。

6. 謝辞

本研究の遂行に当たり、西井龍映九州大学数理学研究院教授(客員研究官)、永田晃也九州大学経済学研究院准教授(前総括主任研究官)、桑原輝隆所長から、多くの御助言・御意見をいただいた。ここに謝意を表する。

また、本研究の遂行に御協力いただいた廣潤社にもお礼を申し上げる。

また、筆者が国立試験研究機関所属であるため、本研究で構築・使用するHM-DBは「行政機関の保有する個人情報保護に関する法律」等の規定の適用を受ける。本研究者は当該法令規定を遵守して研究を実施する。

なお、本研究における主張等の責任は専ら筆者が負うものであり、以上の方々には及ばないことを付け加える。

また、本研究は科学研究費補助金(19710139)の助成を受

けたものである。

附録

第一群:G1(論文シェア 5%～, 4 大学):東北大学, 東京大学, 大阪大学(大阪外国語大学含む), 京都大学

第二群:G2(論文シェア 1～5%, 10 大学):東京工業大学, 名古屋大学, 北海道大学, 九州大学(九州芸術工科大学含む), 筑波大学(図書館情報大学含む), 千葉大学, 神戸大学(神戸商船大学含む), 広島大学, 金沢大学, 岡山大学

第三群:G3(論文シェア 0.5～1%, 17 大学):名古屋工業大学, 東京農工大学, 静岡大学, 三重大学, 信州大学, 富山大学(富山医科薬科大学含む), 山口大学, 岐阜大学, 新潟大学, 東京医科歯科大学, 横浜国立大学, 熊本大学, 群馬大学, 長崎大学, 山形大学, 愛媛大学, 鹿児島大学

第四群:G4(論文シェア 0.5%以下, 55 大学):お茶の水女子大学, 愛知教育大学, 旭川医科大学, 一橋大学, 茨城大学, 宇都宮大学, 岩手大学, 宮崎大学(宮崎医科大学含む), 宮城教育大学, 京都教育大学, 京都工芸繊維大学, 九州工業大学, 弘前大学, 香川大学(香川医科大学含む), 高知大学(高知医科大学含む), 佐賀大学(佐賀医科大学含む), 埼玉大学, 山梨大学(山梨医科大学含む), 滋賀医科大学, 滋賀大学, 鹿屋体育大学, 室蘭工業大学, 秋田大学, 小樽商科大学, 上越教育大学, 政策研究大学院大学, 総合研究大学院大学, 帯広畜産大学, 大阪教育大学, 大分大学(大分医科大学含む), 筑波技術大学(筑波技術短期大学を含む), 長岡技術科学大学, 鳥取大学, 電気通信大学, 島根大学(島根医科大学含む), 東京海洋大学(東京水産・商船大学含む), 東京外国語大学, 東京学芸大学, 東京芸術大学, 徳島大学, 奈良教育大学, 奈良女子大学, 奈良先端科学技術大学院大学, 浜松医科大学, 福井大学(福井医科大学含む), 福岡教育大学, 福島大学, 兵庫教育大学, 豊橋技術科学大学, 北海道教育大学, 北見工業大学, 北陸先端科学技術大学院大学, 鳴門教育大学, 琉球大学, 和歌山大学

参考文献:

- [1] 廣潤社, 全国大学職員録(1988, 1991, 1994, 1997, 2000, 2003, 2005, 2006 年版)。
- [2] 細坪護拳, 大学教員人材の流動性に関する予備的調査, 研究・技術計画学会第 22 回年次学術大会講演要旨, 2G25(2007)。
- [3] 細坪護拳, 大学教員の流動性究明に向けた統計学的アプローチ, 研究・技術計画学会第 24 回年次学術大会講演要旨, 2G01(2009)。
- [4] 細坪護拳, 国立大学教員の流動性計測のためのデータベースの構築とこれによる予備的調査分析, 日本の大学に関するシステム分析-日英の大学の研究活動の定量的比較分析と研究環境(特に, 研究時間, 研究支援)の分析-, 文部科学省科学技術政策研究所, 191-204(2009)。
- [5] (財)文教協会, 文部科学省・国立大学法人等職員録(2008 年版)。
- [6] 日本の大学に関するシステム分析-日英の大学の研究活動の定量的比較分析と研究環境(特に, 研究時間, 研究支援)の分析-, 文部科学省科学技術政策研究所, 27(2009)。
- [7] 平成 18 年度大学等における科学技術・学術活動実態調査報告, 文部科学省科学技術政策研究所, 25-26(2006)。
- [8] 石橋英二, 大学等における科学技術・学術活動実態調査報告(大学実態調査 2007), 文部科学省科学技術政策研究所, 50(2007)。
- [9] 太郎丸博, 人文・社会科学のためのカテゴリー・データ解析入門, ナカニシヤ出版, 163-164(2006)。
- [10] B.S.エヴェリット, 質的データの解析-カイ二乗検定とその展開, 新曜社, 49(1980)。
- [11] アプトン, 調査分類データの解析法, 朝倉書店, 100-102(1980)。
- [12] Alan Agresti, カテゴリーカルデータ解析入門, サイエンス社, 208-210, 224-225(2003)。
- [13] 竹村彰通, 個票開示問題の研究の現状と課題, 統計数理, 51(2), 252(2003)。